Tampere University

EKATERINA KARJALAINEN
(née Olshannikova)

Serendipity and Diversity in Professional Social Matching

Towards diversity-enhancing recommendation strategies

Tampere University Dissertations 909

Tampere University Dissertations 909

EKATERINA KARJALAINEN (née Olshannikova)

Serendipity and Diversity in Professional Social Matching

Towards diversity-enhancing recommendation strategies

ACADEMIC DISSERTATION
To be presented, with the permission of
the Faculty of Information Technology and Communication Sciences
of Tampere University,
for public discussion in the Paavo Koli auditorium
of the Pinni A building, Kanslerinrinne 1, Tampere,
on 15 December 2023, at 13 o'clock.

ACADEMIC DISSERTATION
Tampere University, Faculty of Information Technology and Communication Sciences
Finland

Responsible supervisor and Custos	Professor Thomas Olsson Tampere University Finland	
Co-Supervisor	Senior Research Fellow Harri Siirtola Tampere University Finland	
Pre-examiners	Professor Loren Terveen University of Minnesota United States of America	Assistant Professor Douglas Zytko Oakland University United States of America
Opponent	Senior University Lecturer Antti Salovaara Aalto University Finland	

Cover design: Roihu Inc.

ISBN 978-952-03-3167-2 (print)
ISBN 978-952-03-3168-9 (pdf)
ISSN 2489-9860 (print)
ISSN 2490-0028 (pdf)
http://urn.fi/URN:ISBN:978-952-03-3168-9

Carbon dioxide emissions from printing Tampere University dissertations have been compensated.

PunaMusta Oy – Yliopistopaino
Joensuu 2023

Dedicated to those brave researchers who are open to serendipity and diversity in their professional and personal lives.

PREFACE

This research was carried out at Tampere University from 2017 to 2023 under the academic projects COBWEB and Big Match.

First, I would like to express my deepest gratitude and appreciation to my supervisors, Prof. Thomas Olsson and Dr. Harri Siirtola. It is with immense gratitude that I acknowledge the invaluable support and guidance they provided throughout the research journey. Their mentorship and encouragement have been instrumental in shaping this dissertation.

I would like to express my gratitude to all the members of the research projects. Their willingness to participate and share their experiences has significantly contributed to the findings of this study. Their involvement has been invaluable, and I deeply appreciate their time and contributions. I am especially grateful to my coauthors: Dr. Jukka Huhtamäki, Susanna Paasovaara, Prof. Hannu Kärkkäinen, Prof. Henri Pirkkalainen, Peng Yao, Dr. Erjon Skenderi, and Sami Koivunen.

I also extend my heartfelt thanks to all the members of the Technology × Social Interaction research group (TSI) and the Tampere Unit for Computer-Human Interaction (TAUCHI). Their insightful discussions, constructive feedback, and friendship have greatly enriched my research experience. It has been truly inspiring to collaborate with such talented individuals.

I would like to express my deepest appreciation to all those who have contributed to my academic and personal growth. I am truly fortunate to have such remarkable individuals in my life, and I am indebted to every one of them for their contributions. I extend special gratitude to Prof. Yevgeni Koucheryavy, Prof. Sergey Andreev, and Dr. Aleksandr Ometov, who opened the door to the exciting world of academia for me.

I am immensely grateful to my friends and family, whose unwavering support, understanding, and encouragement have been my pillar of strength throughout this journey. I want to acknowledge my best friends, Dr. Margarita Gapeyenko, Alisa

Burova, and Chelsea Kelling, for supporting me throughout this journey. Their belief in me has fueled my determination to overcome challenges and persevere in my research endeavors. Their love and encouragement have been my constant motivation.

This dissertation is a culmination of collective efforts, guidance, and support from peers at the TSI research group, the TAUCHI Research Center, colleagues, friends, and family. I am forever grateful for their involvement and their belief in my abilities. Finally, I would like to express my gratitude to all the study participants. Without their contributions, this work would not have been possible.

ABSTRACT

Professional Social Matching (PSM) is the practice of building and maintaining connections in the context of knowledge work. Various people recommender systems and social matching applications have been designed to facilitate PSM by finding relevant others among numerous options. However, conventional recommendation approaches have been found to support algorithmic and human biases, disrupting knowledge flow and social networking, which is vital for PSM. This dissertation focuses on two central concepts: diversity and serendipity. Diversity refers to the importance of exposing individuals to different perspectives, backgrounds, and experiences to foster productive and creative knowledge work. Serendipity, on the other hand, pertains to the occurrence of unsought yet valuable connections that can lead to unexpected and fortunate encounters.

The research questions driving this dissertation revolve around the role of diversity and serendipity in PSM tools and the manifestation of these concepts in recommendation strategies. The research process involved a series of five publications. The first two publications employed online surveys to investigate social serendipity and the processes in making valuable connections in online and offline realms. The third publication entails a literature review with a specific emphasis on the conceptual framework of Big Social Data (BSD), as its comprehension holds significant relevance for the domain of user modeling within recommender systems. The last two publications experimented with diversity-enhancing recommendation strategies and examined the alignment between subjective perceptions and objective measures of recommendation relevance.

The findings uncovered diverse insights into the characteristics and antecedents of social serendipity, highlighting the necessity for identifying novel mechanisms to foster serendipity experiences in PSM. The results also revealed consistent and significant differences in subjective perceptions of the proposed diversity-enhancing strategies, thus indicating their preliminary effectiveness. Participants showcased the

ability to identify relevant others at all levels of similarity and structural network positions, despite the inherent bias in selection. The research contributions lie in elucidating the proactive and reciprocal sense-making involved in PSM, identifying qualities that foster serendipitous encounters, exploring the potential of Big Social Data, and developing and evaluating recommendation mechanisms that promote diversity in professional social networks.

CONTENTS

List of Figures

List of Tables

ABBREVIATIONS

BSD	Big Social Data
CSCW	Computer-Supported Cooperative Work
GDPR	General Data Protection Regulation
HCI	Human-Computer Interaction
IR	Information Retrieval
LDA	Latent Dirichlet Allocation
NLP	Natural Language Processing
PSM	Professional Social Matching
RecSys	Recommender Systems
RtD	Research Through Design
TF-IDF	Term Frequency-Inverse Document Frequency

ORIGINAL PUBLICATIONS

Publication I	Olshannikova, E., Olsson, T., Huhtamäki, J., Paasovaara, S., & Kärkkäinen, H. (2020a). From chance to serendipity: Knowledge workers' experiences of serendipitous social encounters. *Advances in Human-Computer Interaction*, *2020*(1827107), 1–18. https://doi.org/10.1155/2020/1827107
Publication II	Olshannikova, E., Pirkkalainen, H., Olsson, T., & Huhtamäki, J. (2022a). What supports serendipity on twitter? online survey on the role of technology characteristics and their use. *25th International Academic Mindtrek conference (Academic Mindtrek 2022), November 16–18, 2022, Tampere, Finland*, 13. https://doi.org/10.1145/3569219.3569346
Publication III	Olshannikova, E., Olsson, T., Huhtamäki, J., & Kärkkäinen, H. (2017). Conceptualizing big social data. *Journal of Big Data*, *4*(3), 1–19. https://doi.org/10.1186/s40537-017-0063-x
Publication IV	Olshannikova, E., Olsson, T., Huhtamäki, J., & Yao, P. (2019a). Scholars' perceptions of relevance in bibliography-based people recommender system. *Computer Supported Cooperative Work (CSCW)*, *28*(3–4), 357–389. https://doi.org/10.1007/s10606-019-09349-w
Publication V	Olshannikova, E., Skenderi, E., Olsson, T., Koivunen, S., & Huhtamäki, J. (2022b). Utilizing structural network positions to diversify people recommendations on twitter. *Advances in Human-Computer Interaction*, *2022*(6584394), 17. https://doi.org/10.1155/2022/6584394

Author's contribution

Publication I	Olshannikova planned the study in collaboration with the authors. She conducted the study and gathered and analyzed the data. Olsson, Paasovaara, and Olshannikova performed an iterative analysis of the collected qualitative data. She was the principal author and oversaw the publication's production.
Publication II	The authors planned and designed the study collaboratively. Olshannikova collected the data. Pirkkalainen primarily performed the statistical data analysis. Olshannikova developed the article's framework and focus with the authors. She was the principal author and oversaw the publication's production.
Publication III	Olshannikova introduced the literature review topic to the other authors and coordinated the process to complete the manuscript. The authors worked together to develop the article's framework and focus. Olshannikova performed the primary literature review and analysis. She was the principal author and oversaw the publication's production.
Publication IV	Olsson, Huhtamäki, and Olshannikova planned and designed the experiment. Yao implemented the recommendation system, including bibliographic data collection, cleaning, processing, and generating recommendations. Olshannikova conducted the controlled experiment, collected and analyzed the data on the users' subjective perceptions of received recommendations. She was the principal author and oversaw the publication's production.
Publication V	The authors planned and designed the controlled experiment collaboratively. Skenderi implemented the recommendation system, including Twitter data collection, cleaning, processing, and generating recommendations. Koivunen and Olshannikova recruited participants, conducted the experiment, and collected the data on subjective perceptions of recommendations. Olshannikova analyzed and reported the data and was the principal author overseeing the publication's production.

1 INTRODUCTION

Professional Social Matching (PSM) (Olsson et al., 2020) is the practice of building and maintaining connections in the context of knowledge work, i.e., occupations that require extensive intellectual skills to create, distribute, or apply knowledge (Davenport, 2005). PSM is ubiquitous in knowledge work, which is highly dependent on collaboration and social networks. The structure and quality of individual social networks can influence information flow, opinion expression and discourse, exchange of ideas, and production of new knowledge. Historically, people have expanded their networks through traditional face-to-face socialization practices (e.g., professional social events), which can be demanding and time-consuming. This has become more manageable with the rise of social media, which exposes diverse communities to one another despite spatial or temporal boundaries (Muller et al., 2012). However, limitless social networking opportunities also create the paradox of choice: How do you find relevant people in an enormous pool of potential connections?

Facilitating PSM has become a central goal in information and communication technology in both the commercial and academic sectors. This has resulted in various people recommender systems (Beel et al., 2016; Guy & Pizzato, 2016) and social matching applications (J. M. Mayer et al., 2016), which orchestrate a way to build and maintain new connections. Social media networks have also contributed to computational social matching by introducing new contact suggestions or features like "whom to follow." Such services typically use diverse social data to model similarities or differences between users to connect them to relevant others (Kunaver & Požrl, 2017). However, most current approaches to recommender systems and social network analysis in the context of PSM demonstrate the threat of long-term detrimental implications for collaboration practices and knowledge work. For example, conventional mechanisms, such as optimizing for similarity and triadic closure (Carullo et al., 2015), involve the risks of strengthening homophily bias (Kossinets & Watts, 2009; McPherson et al., 2001) and echo chambering (C. T. Nguyen, 2020; Terren

& Borge-Bravo, 2021).

The similarity-maximizing approach has been adopted from item recommender systems, in which similarity metrics were treated as a proxy of relevance (Guy et al., 2010; Heck, 2013). However, recommending people is fundamentally different from recommending items because the goal is to establish reciprocal relationships (Koprinska & Yacef, 2015); the success of the recommendation depends on whether both the user and the recommendee follow up to establish a connection. As objects of recommendation, individuals contain many diverse characteristics that influence decision-making when evaluating the usefulness of potential connections, especially in a professional context (Olsson et al., 2020). Although collaboration within a group of people with shared interests can contribute to a safe and trustworthy work environment, diversity exposure is essential for creating new knowledge (Y. C. Yuan & Gay, 2006). Even so, recommending people based only on similarity or diversity does not imply that the recommendation would be perceived as relevant. In that regard, one dimension that, by definition, adds value is serendipity (Adamopoulos & Tuzhilin, 2014), which manifests as unsought, surprising, and fortunate encounters. Serendipity is a worthwhile goal in establishing new relationships to achieve more productive and creative knowledge work and collaboration (Jarrahi, 2017; Olshannikova et al., 2020a; Parise et al., 2015). Although diversity and serendipity were studied in various disciplines (Kaminskas & Bridge, 2016; Muller et al., 2012), their factual value for PSM and how they should manifest in recommendation strategies remain unclear and require more research.

Therefore, this dissertation's primary goal is to investigate how the concepts of diversity and serendipity can unfold in the design of PSM tools. The thesis is compounded by five publications that subscribe to the idea of diversity exposure, especially when considering social networking for professional purposes, focusing on one-to-one social matching in knowledge work. The first two articles (online surveys) investigate how social serendipity, as a manifestation of successful PSM, unfolds in the processes of making unsought yet valuable connections in both the online and offline realms. The third publication is a literature review that explores and conceptualizes Big Social Data (BSD), aiming to grasp the magnitude and potential of rich data. BSD serves as an enabler of user modeling and social recommendation as its analysis provides a deeper understanding of users' interests, their network dynamics, and how they engage within their communities. The last two publications experiment with diversity-enhancing recommendation strategies and demonstrate how to derive user qualities from users' self-representation, actions, and interactions online to deliver professionally relevant people recommendations. All publications except the third involve collecting and analyzing subjective data (interview or questionnaire data). The last two papers also compare subjective user perceptions with objective measures of the recommendations' relevance.

In the following, I first provide a conceptual overview of diversity and serendipity as key concepts of the dissertation. Then, I proceed by addressing the scope and focus of this work. The research questions, objectives, and contributions are described, followed by a report on the research process and methods used. Finally, the structure of the thesis is presented.

1.1 Key Concepts—Diversity and Serendipity

Research on diversity and serendipity has resulted in different conceptualization perspectives (see Figure 1.1). For example, in organization studies, these concepts are treated as behavioral and social patterns worth pursuing due to their positive effects on knowledge work and collaboration (Jarrahi, 2017; Mitchell & Nicholas, 2006; Muller et al., 2012; Parise et al., 2015). In computer science, both diversity and serendipity have been actively studied in the field of Information Retrieval (IR) (Kaminskas & Bridge, 2016). IR research first acknowledged the detrimental effect of using maximal similarity to achieve accuracy in information discoveries and addressed the need for diversification (Agrawal et al., 2009; Carbonell & Goldstein, 1998; Clarke et al., 2008). Later, this idea was adopted in Recommender Systems (RecSys), which addressed the need to overcome algorithmic biases (Edizel et al., 2020) and personalization drawbacks in information filtering (Koene et al., 2015; Pariser, 2011; Sîrbu et al., 2019). Thus, diversity and serendipity have been treated as recommendation dimensions and approaches to increase user satisfaction (Smyth & McClave, 2001; Ziegler et al., 2005).

Human-Computer Interaction (HCI) research is focused on the perceived diversity of social relationships, exploration of diversity dimensions (e.g., cognitive, physiological, and demographic differences), and translation of gathered insights into guidelines establishing fairer technology designs (Himmelsbach et al., 2019; Robert, 2016). Computer-Supported Cooperative Work (CSCW) research investigated the

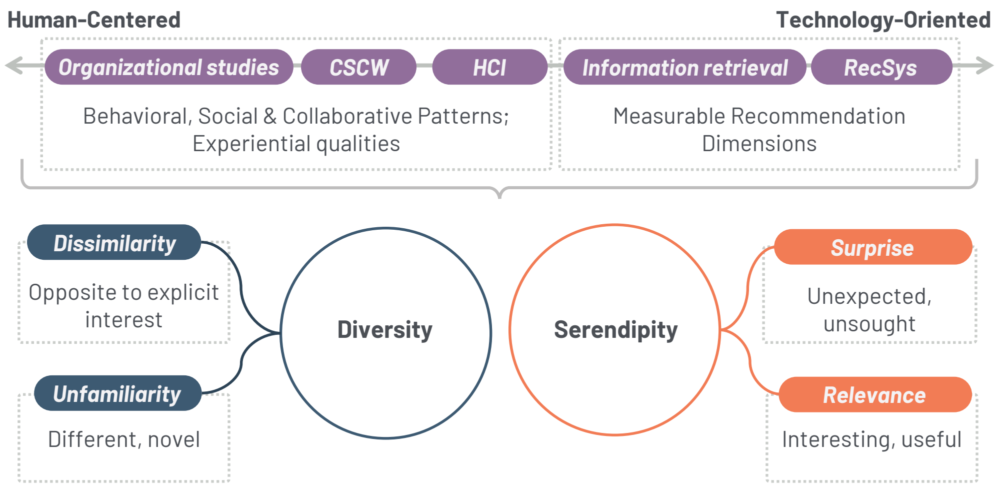


Figure 1.1 Visual interpretation of the diversity and serendipity conceptualizations based on prior research.

role of differences in individual qualities (e.g., cultural diversity) in collaboration practices (Dong et al., 2016; Wang et al., 2011), addressed strategies for diversity exposure, and designed solutions to overcome the challenge of filter bubbles (Ookalkar et al., 2019; Resnick et al., 2013). Regarding serendipity, HCI and CSCW research explored the design space for closely related topics such as opportunistic social matching (J. Mayer, 2014; J. Mayer & Jones, 2016; J. M. Mayer et al., 2015; Paasovaara et al., 2016), increasing social awareness, and allowing chance encounters (Eagle & Pentland, 2005; Erickson & Kellogg, 2000; Jeffrey & McGrath, 2000).

There are different abstraction levels for conceptualizing and defining serendipity and diversity. The first human-centered conceptualization approaches diversity and serendipity as experiential qualities and focuses on subjective perceptions of these qualities and the societal effects they bring. The second technology-oriented conceptualization refers to diversity and serendipity as measurable dimensions that can be algorithmically designed and evaluated. The common conceptual aspect in the prior literature is that diversity is seen as the opposite of similarity. RecSys research, for instance, defines it as average dissimilarity (Bradley & Smyth, 2001), distributional inequality (Fleder & Hosanagar, 2007), and non-redundancy (Vargas et al., 2014). Therefore, it can be interpreted as a perceived difference or a measurable distance between all recommended items (presented to the user). A concept closely related to diversity is novelty, a recommendation that is unknown by the user or objectively different from previously recommended items (Vargas & Castells, 2011). While diversity and novelty are distinguished in the prior literature, this dissertation combines their qualities under one definition. Thus, in this dissertation, the following conceptualization is used:

> ***Diversity*** *is conceptualized as encounters with something dissimilar (the opposite of users' explicit interests) and unfamiliar (different and novel).*

While diversity does not necessitate relevancy, serendipity, by definition, is an experience that results in a valuable outcome (McCay-Peet & Toms, 2015). Common elements of serendipity in the literature include surprise and relevance (Kotkov et al., 2016). In RecSys, serendipity is primarily defined as a difference in recommended items in relation to users' expectations (Kaminskas & Bridge, 2014). Some RecSys research also suggests that the surprise element in serendipity makes encounters novel by definition (new) (Iaquinta et al., 2010; McNee et al., 2006). However, serendipity can occur when encountering others who are familiar or unfamiliar (Adamopoulos & Tuzhilin, 2014), and the experience will depend on contextual settings and how the encounters satisfy the individuals' needs at the moment (McCay-Peet & Toms, 2015). In this work, the notion of novelty is excluded from the conceptualization:

> ***Serendipity*** *refers to encounters with something surprising (unsought, unexpected) and relevant (valuable, useful) that results in a positive emotional response.*

Although there have been attempts to replicate the elements of serendipity in recommender systems, only users can judge whether a recommendation is perceived as serendipitous.

1.2 Scope, Focus and Empirical context

The dissertation draws on the conceptualization of PSM (Olsson et al., 2020), focusing on a one-to-one social matching scenario within knowledge work as an empirical context. This work excludes social matching for romantic purposes, as the context of dating has received significantly more research attention and comprehensive studies compared to the domain of PSM. Additionally, the matching criteria and goals differ significantly in the dating and professional contexts. For example, PSM involves criteria such as skills, expertise, industry experience, career goals, education, and professional interests to ensure a beneficial and synergistic relationship. Romantic matching, in turn, more personal and intimate communication, expressions of affection, romantic gestures, dates, and shared experiences to foster emotional closeness. Therefore, limiting the focus to the professional context concentrates investigative efforts exclusively on the intricacies and dynamics inherent to PSM, unencumbered by the complexities and multifaceted nature of other social domains. Consequently, it allows to establish precise research objectives and formulate targeted research questions.

PSM has been studied by various research fields, which defines the multidisciplinary nature of the dissertation. This research is positioned at the intersection of the following fields (See Figure 1.2):

- *HCI*, which focuses on the user-centered design and evaluation of technology (social matching applications, in this context);
- *CSCW*, which studies computational collaboration tools and techniques as well as their psychological, social, and organizational effects;
- *RecSys*, which concentrates on user modeling, context, and social networking analysis in the design and evaluation of people-to-people recommendation strategies.

From HCI and CSCW perspective, this dissertation aims to understand the behavioral, social, and contextual factors in one-to-one social matching and how they inform the design space for ICT-facilitated PSM. From a RecSys perspective, this work addresses the design and evaluation of diversity-enhancing people-to-people recommendation strategies for PSM. The HCI field informs the dissertation's focus on human-centered evaluation, concentrating on subjective perceptions of recommendations' relevance rather than the efficiency and accuracy of people-recommender artifacts.

1.3 Research Questions, Objectives, and Contribution

The dissertation's goal is to explore the design space for serendipity-inducing and diversity-enhancing mechanisms for PSM. The goal is achieved by addressing the following research questions:

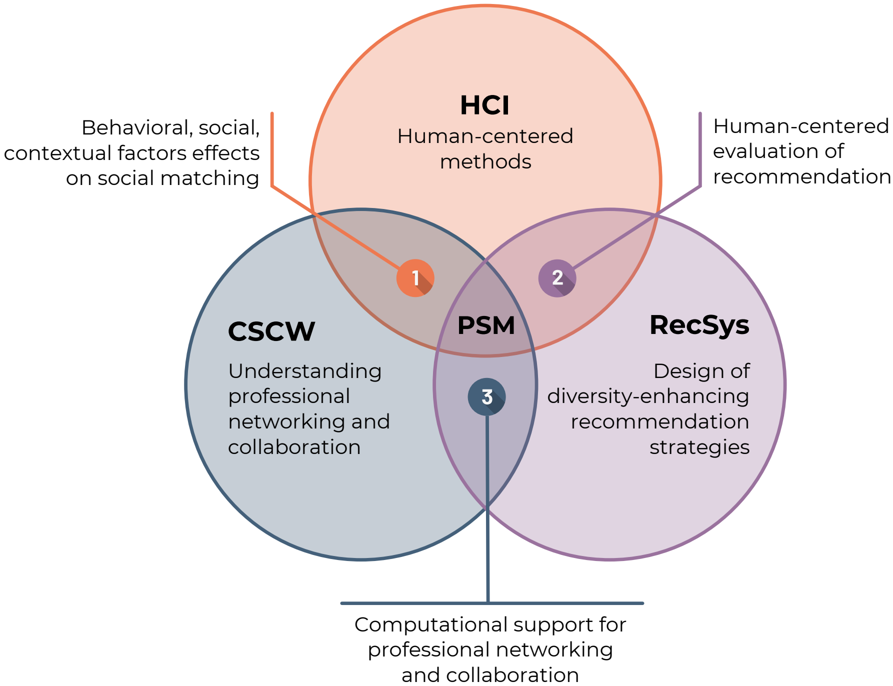


Figure 1.2 Scope, focus, and empirical context of the dissertation.

- *RQ1. How can experiences of social serendipity inform the design of diversity-enhancing people recommendation strategies?* This question investigates subjective experiences of social serendipity as manifestations of successful PSM. The research question is addressed by the objectives of Publications I and II (see Table 1.1). Understanding the characteristics of social serendipity and related factors can inform design directions for diversity-enhancing people recommendation strategies.
- *RQ2. How can diversity-enhancing recommender strategies facilitate professional social matching?* This research question implies the implementation and application of diversity-enhancing strategies in people recommendation systems for PSM. It seeks approaches to model users based on social data and uses diversity-inducing clustering in the recommendation pool, including content and social networking analyses. This research question is addressed by the objectives of Publications III, IV, and V (see Table 1.1).
- *RQ3. How do users subjectively perceive recommendations based on diversity-*

enhancing strategies? This research question implies the evaluation of the proposed diversity-enhancing strategies. Following a human-centered approach, this question seeks to conceptualize measures of people-recommendation relevancy for the user rather than predict the accuracy and effectiveness of the recommendation algorithms. This work focuses on designing human-centered experiments that enable the collection of objective and subjective observations. This question is addressed by the objectives of Publications IV and V (see Table 1.1).

The contribution and novelty of the research constitute the following. First, the dissertation provides empirical insights into the characteristics and antecedents of social serendipity, highlighting the necessity for identifying novel mechanisms to foster serendipity experiences in PSM. Second, this work introduces diversity-enhancing strategies for people recommenders in PSM, which is an emergent research topic related to the HCI, CSCW, and RecSys fields. Third, the dissertation proposes a socially acceptable way to use BSD to derive insight into individuals' personal qualities for social matching. Fourth, this study contributes to the user-centric evaluation of people recommender systems. Operationalized measures of subjective perceptions can be utilized in future research to evaluate social relevance. Overall, the dissertation's findings contribute to the design of people recommender systems. This work also opens new research directions and design guidelines for unconventional social matching systems calibrated according to the similarity-diversity continuum and serendipity.

1.4 Research Process and Methods

The research follows a human-centered design process with two phases (see Figure 1.3). The first phase is exploratory and dedicated to defining enablers for people recommenders—BSD and social serendipity. This phase contributed to RQ1 and RQ2 by building theoretical grounds and resulted in:

- A literature review on the conceptualization of data types that represent online self-representation, actions, and interactions;
- Two online surveys on social serendipity to derive subjective perceptions of successful PSM, social relevance factors, and serendipity antecedents.

Table 1.1 Objectives of publications

Publication	Objectives
Publication I	• To provide a detailed qualitative account of the various experiential and contextual qualities in social serendipity; • To define key characteristics of social serendipity; • To discuss the role of technology in social serendipity.
Publication II	• To investigate different uses of Twitter and the characteristics of technology as antecedents of work-related serendipity; • To explore the role of personality characteristics in experiencing serendipity on Twitter.
Publication III	• To conceptualize and define BSD; • To provide classification of BSD types that are available for research and analysis; • To provide a research agenda for BSD-driven future works and potential implications for practice and research.
Publication IV	• To implement diversity-enhancing mechanism based on calibration of similarity-difference continuum of the content; • To study how academics perceive relevance, complementarity, and diversity of individuals in their profession and how these concepts can be embedded optimally in people recommenders.
Publication V	• To implement diversity-enhancing mechanism based on different structural network positions on Twitter; • To extend the discussion on matching strategies by utilizing analytical mechanisms beyond content-based similarity and triadic closure; • To obtain subjective perceptions on relevance of recommendations.

The second phase aimed to design and evaluate diversity-enhancing strategies for people recommender systems in practice. This phase contributes primarily to RQ2 and RQ3, resulting in empirical findings from two controlled experiments and operationalization of evaluation measures for people recommenders.

This dissertation applies a Research Through Design (RtD) approach (Dautenhahn & Ghauoi, 2014), in which design is part of the research and plays an essential role in contributing to the generation of new knowledge. In HCI, RtD discovers,

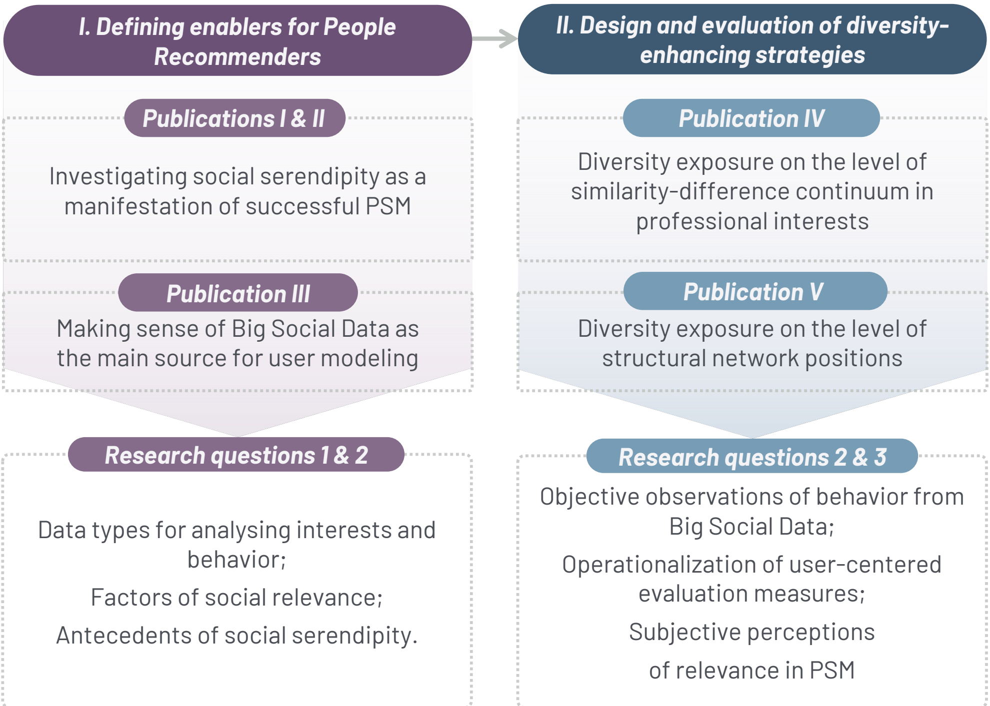


Figure 1.3 Research process consisting of two main phases.

exemplifies, clarifies, and promotes principles that can bring about social change, which manifests in this dissertation as a demonstration of the need for diversification in PSM. Throughout the process, multiple research methods were used (see Table 1.2).

The first phase of the dissertation consists of research to generate knowledge that translates into actionable design directions for the second phase, the development of diversity-enhancing people recommendation artifacts. The artifacts, in turn, make interactions with diversity-enhancing people recommenders and subjective perceptions observable through design. Conducting experimental research with developed people recommender artifacts exemplifies whether it is possible to overcome algorithmic and human biases in PSM and results in the production of new knowledge.

1.5 Research Ethics

In conducting this research, utmost consideration was given to ethical principles and guidelines to ensure the protection and well-being of all participants involved. We

Table 1.2 Research methods used to address research questions.

	Research Approach	Data Gathering	Analysis
RQ1	International online surveys; scenarios	Likert-scale statements and open-ended questions	Qualitative data analysis – structural, axial, and focused coding in Nvivo; Descriptive Statistics in Tableau; Factor analysis and Linear Regression in SPSS
RQ2	Scoping, critical and narrative review	Selective search based on keywords	Thematic analysis; Critical interpretive synthesis; Conceptual frameworks
	Content and social network analyses	DBLP database; Twitter API	Natural Language Processing (NLP), Cosine Distance, Term Frequency-Inverse Document Frequency (TF-IDF), Latent Dirichlet Allocation (LDA), Louvain Modularity algorithm
RQ3	Controlled experiments, semi-structured interview, international online survey	Likert-scale questionnaires and open-ended questions	Qualitative data – elemental, axial, and focused coding for qualitative data; Descriptive statistics in Tableau; Friedman test, Spearman Correlation test in RStudio and SPSS

followed the policies provided by the National Ethical Committee in Finland. Accordingly, the studies did not require an ethical review, as they included informed consent and did not involve any of the following: underage subjects, exposure to strong stimuli, potential long-term mental distress, or intervention with the participants' physical integrity. The studies were identified as low-risk and, therefore, did not require an ethical review. The following summarizes the ethical considerations:

- *Informed Consent.* Informed consent was obtained from all participants before their involvement in the study. Consent comprehensively explained the studies' objectives, procedures, potential risks, and benefits. The participants were assured of their right to voluntary participation, their ability to withdraw from the study at any time without penalty, and the confidentiality and anonymity of their responses.
- *Confidentiality and Anonymity.* To ensure confidentiality, all data collected during the study were anonymized and stored securely. Each participant was assigned a unique identifier code, and personal identifying information was

erased from the research data. Only authorized researchers (co-authors) had access to the data.

- *Data Analysis and Reporting.* The confidentiality of the participants' responses was maintained, and no identifiable information was disclosed in the final report or any subsequent publications.
- *Dissemination of Findings.* When disseminating the research findings, efforts were made to ensure the results were presented responsibly and transparently. Additionally, participants were acknowledged for their valuable contributions while their anonymity was maintained.

By upholding these ethical considerations, this research aims to contribute responsibly and ethically to the body of knowledge.

1.6 Structure of the Thesis

The thesis consists of five chapters. The Introduction (**Chapter 1**) describes the background and motivation for the dissertation and addresses the key concepts of diversity and serendipity. The Introduction also describes the scope, focus, and empirical context and highlights the research questions, objectives, and contributions, followed by a description of the research process and methods.

Chapters 2-4 have a similar structure and synthesize findings from publications under the dissertation's three main topics: social serendipity (*Publication I & II*), the design of diversity-enhancing strategies for people recommendations (*Publications III, IV & V*), and the evaluation of people recommender systems (*Publications IV & V*). These chapters explain the related research gaps and continue with the explanation of the related research and findings conducted in this dissertation.

Accordingly, **Chapter 2** starts with a review of the fundamental motivational factors and human needs in strategic PSM. Then, I discuss prior literature on computational support for PSM. The chapter introduces the definition and conceptualization of social serendipity, providing an account of relevant factors for unexpected yet beneficial professional encounters. **Chapter 3** provides an overview of existing applications and contexts of use for people recommender systems, as well as recommendation approaches. **Chapter 3** also describes the proposed diversity-enhancing mechanisms. **Chapter 4** is dedicated to evaluating people recommender systems, starting with a literature review on system-centric and user-centric evaluation approaches and the concept of the user experience in recommender systems. It provides an overview of custom measures for collecting subjective perceptions on recommendations generated based on proposed diversity-enhancing strategies, summarizing findings from *Publications IV & V*.

The Discussion (**Chapter 5**) provides a reflection and conclusions to the findings. First, the chapter revisits the research questions and the dissertation's contributions. Then, it addresses design considerations and limitations and outlines directions for future work.

2 UNDERSTANDING SOCIAL SERENDIPITY

This chapter synthesizes the results of two empirical studies on social serendipity (Publications I & II; see the summary in Figure 2.1). While both studies investigate the topic in professional contexts, they address different research questions and apply different methods. The first study focuses on qualitatively understanding lived experiences of social serendipity, and the second focuses on quantitatively understanding the role of technology and other antecedents in the emergence of social serendipity. In the following, I first discuss the conceptualization of PSM as a temporal process and exemplify the interest in developing technology to support it. Next, the conceptualization of social serendipity is presented. The key findings from the two studies are then discussed.

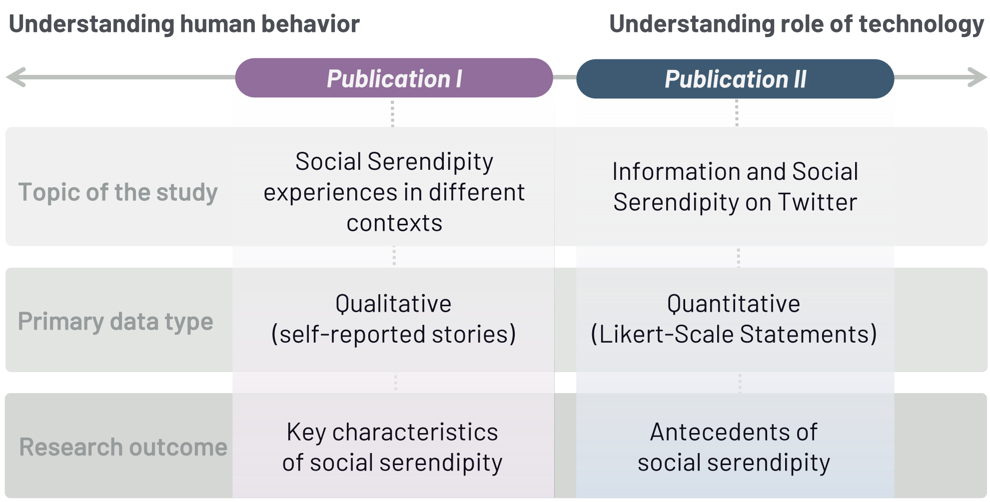


Figure 2.1 Overview of Publication I and Publication II, highlighting differences between studies in relation to research focus, topic, data types, and outcomes.

2.1 Background: Professional Social Matching and Social Serendipity

2.1.1 Strategic Professional Social Matching

PSM is a process in which new ties are established to exchange beneficial interpersonal resources for work-related activities. The term "PSM" was coined relatively recently. Therefore, most prior research covered in this subsection addresses the topic from the perspective of social networking for professional gain. Previous research indicated that strategic social networking positively impacts individual and organizational performance, fostering career success (Collins & Clark, 2003; Cross & Thomas, 2011; Kay, 2010; Sparrowe et al., 2001). Conceptually, it is seen as a goal-directed activity and as a behavior. For example, goal-directed networking aims to improve leadership and work performance and facilitate job tasks (Michael & Yukl, 1993; Yukl, 2012) by accessing interpersonal resources (e.g., valuable information and data). Alternatively, goal-directed networking can be related to acquiring connections to increase visibility, open new opportunities for professional growth (Claes & Ruiz-Quintanilla, 1998; De Vos et al., 2009), or succeed in a job search (Granovetter, 2018; Van Hoye & Saks, 2008). This dissertation focuses on the second conceptualization and approaches PSM as a set of networking behaviors (Porter & Woo, 2015; Wolff et al., 2008) to facilitate work-related activities by maximizing mutual advantages in dyadic relationships.

Porter and Woo (2015) have distinguished three fundamental phases in networking: initiation, growth, and maintenance. Subscribing to Porter and Woo (2015)'s conceptualization, I provide a visual interpretation of the PSM process (see Figure 2.2). In contrast to Porter and Woo (2015), I divide the initiation phase into two parts. First, the connection starts with the *introduction* of two individuals guided by a rational schema (curiosity or interest) about the prospective benefits of establishing a new relationship. External and internal contexts often influence this phase. It requires that at least one of the individuals has a need to establish new ties (e.g., they need to find an expert) and that opportunities are available to approach others (e.g., at a networking event). Following a successful introduction, *initial exchange interactions* may begin, where individuals assess each other's valued resources and deliberate what to offer and ask for in return. If a mutual benefit is revealed, the individuals might proceed to the *growth* phase, which is characterized by building behaviors (reciprocation) to exchange interpersonal resources and establish trustworthiness. If the interactions are mutually rewarding and perceived as trustworthy, the individuals reach the *maintenance* phase, resulting in mature relationships based on integrity and benevolence. At this phase, the individuals are willing to share interpersonal resources, even if it is advantageous only for a single individual.

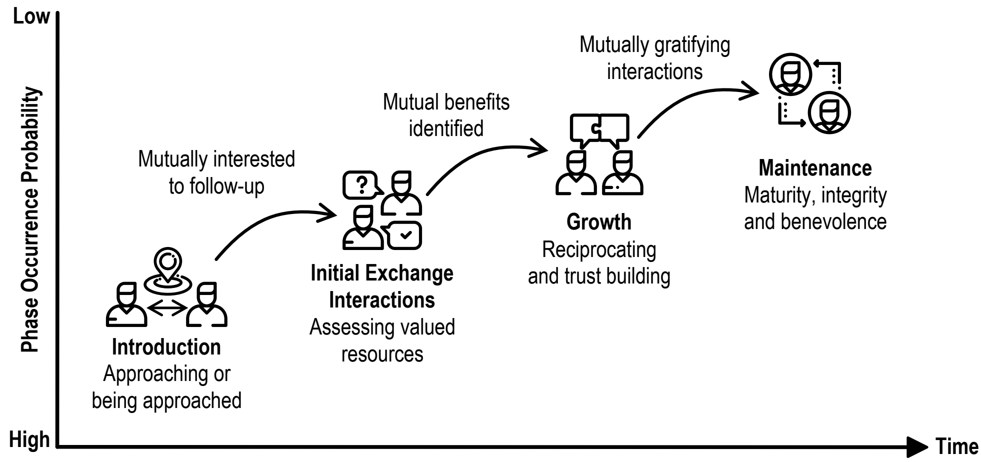


Figure 2.2 Visual interpretation of PSM processes. Occurrence probability decreases for phases that happen later.

PSM progresses over time, and the probability of establishing mature relationships is low. Only a few new encounters reach even the second phase if there is no mutual interest in following up and investigating prospective benefits. Similarly, both parties should identify the mutual advantages to achieve the growth phase. Typically, after valued resources are shared and used for work-related activities, interactions may end and not reach the maintenance phase. The individuals' interactions should be mutually gratifying to achieve maturity, when they begin to value the connection in their social circles and trust that benefits will follow without the need to ask.

2.1.2 Computational Support for Professional Social Matching

Supporting PSM with computational solutions has been a prominent topic for decades in the commercial and academic sectors. LinkedIn is one of the most well-known commercial services, primarily supporting strategic networking for job searches and simplifying recruitment processes. Various applications have been built to facilitate the matching of like-minded professionals, such as Shapr[1], Common Connect App[2], Brella[3], and Invitely[4]. These applications primarily utilize simplistic mechanisms to present social recommendations, such as a profile picture, name, affiliation, and a short bio or keywords of interest (see Figure 2.3). There are also non-commercial event-based services designed to support social networking among scholars and enhance the experience of research communities at conferences, such as Find & Connect (Chin et al., 2014), Confer (A. X. Zhang et al., 2016), Conference Navigator (Brusilovsky et al., 2017).

Figure 2.3 Examples of mechanisms for presenting social recommendations in modern social matching applications for building professional networks.

The contemporary strategy is to introduce social recommendations in a list format that emphasizes the profile picture and provides fewer details about the professional persona (see Figure 2.3). Such approaches have been derived from a content

[1]See https://shapr.co
[2]See https://www.commonconnectapp.com
[3]See https://www.brella.io
[4]See https://invitely.com

recommendation system, where seeing the product or item is essential. Such an approach might be sufficient for the lightweight and low-cost considerations in social matching (e.g., adding a known contact to a friend list). However, PSM is distinguished by the complex spectrum of collaboration and partnering needs, where it is difficult to judge the relevance of potential connections based only on appearance. Prior research acknowledges the limitations and biases that traditional profile presentation can bring to PSM (Archer & Zytko, 2019). The profile picture still plays a central role in the design gaudiness for PSM, however, research suggests utilizing visual assets that depict collaboration and expertise opportunities rather than physical appearances (Zytko & DeVreugd, 2019).

In terms of supporting social interaction, such applications and services mostly facilitate the first phase of the PSM process—introducing opportunities. The rest depends on direct communication between individuals to determine whether there is a mutual match and a willingness to interact. Therefore, there are few encouraging or persuasive features, because these networking services are designed primarily for goal-oriented individuals who can formulate their needs in a simple profile. Although using such services might expand the number of encounters, there is limited support for value creation between professionals. Focusing on the users' similar interests or goals also disregards their need for complementary expertise. However, knowledge workers would benefit from lucky, unexpected encounters and proactive suggestions from the system rather than a user-initiated search for similar others. Learning from examples of social serendipity can facilitate social interactions with the design of non-conventional recommendation strategies.

2.1.3 Social Serendipity: Happystances in Professional Life

Serendipity refers to uncontrolled circumstances that lead to unexpected yet fortunate discoveries (Merton & Barber, 2011) prompted by an individual's interaction with ideas, information, or objects (McCay-Peet & Toms, 2015). Prior research has investigated the phenomenon primarily within the context of information retrieval (Agarwal, 2015; Bawden, 2011) and knowledge building (Buchem, 2011; Nutefall & Ryder, 2010), giving birth to the term *information serendipity*. This type of serendipity results in valuable encounters with content or knowledge, which is advantageous and impactful for creativity and innovation (Anderson, 2011; Johnson, 2011). However, experiences where unsought valuable social connections are made (*social serendipity*) have been acknowledged but not studied empirically (McCay-Peet & Wells, 2017). In professional life, social serendipity can result in fruitful collaboration, successful recruitment, and the discovery of novel information from peers. For example, one of the participants in the first study (Publication I) describes his story:

> *"I was at the ski slope trying to do downhill for my second time. It was a tough day, and I was completely exhausted from falling, hitting other people and trees. A person, who was passing by few times, stopped and started to teach me how to turn and brake. He spent around 20 minutes with me. During the next day, I have met him again but on the lift. We had a nice life deep conversation and spent some time skiing. During our short lift-talk, we ended up realizing that our research interests and fields are almost the same. We have changed cards and now have a collaboration going!"* (Male, 25 y.o., Researcher)

Similar to PSM, serendipitous experiences are often conceptualized as a process-based activity. Various models have been proposed to describe key phases and important factors (Makri & Blandford, 2011; McCay-Peet & Toms, 2015; Pease et al., 2013; Sun et al., 2011). Figure 2.4 shows the key phases of social serendipity mapped to the PSM process. In PSM processes, first the encounter is initiated. In serendipity experiences, the individuals' introduction occurs due to so-called *Triggers*, which in the present research are broken into *Contextual* (environmental settings) and *Interaction* (social settings). For example, contextual triggers refer to places where the encounter happened (in the example story, a ski slope); interaction triggers refer to initial social engagement (e.g., bumping into a person several times leads to an initial conversation). The assessment of instrumentality in social serendipity occurs within the *Connection* phase—the proactive clarifying of triggers and estimating prospective benefits. For instance, this phase relates to the revelation of commonalities (similar interests, experiences) or complementary skills beneficial to both individuals. The deliberate assessment of valued interpersonal resources can lead to the growth of social relationships and motivate individuals for *Follow-up* activities (e.g., repeated meetings) to reach a *Valuable Outcome* (work-related or positive personal impacts). Therefore, the valuable outcome phase refers to the subjective value of newly established relationships that motivate people to advance and maintain their connection.

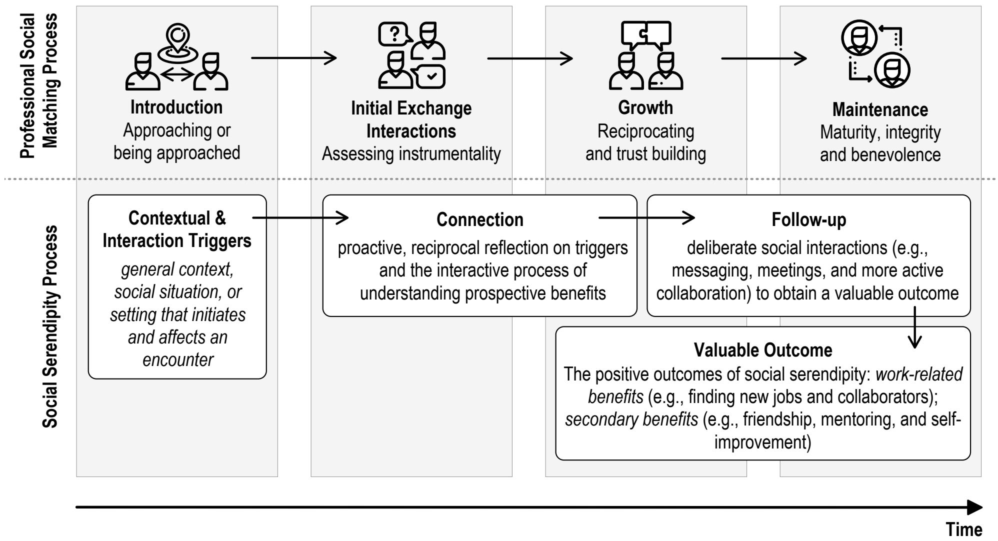


Figure 2.4 Social serendipity process in relation to PSM phases.

Although the general notion of PSM does not require progressing relationships to reach a valuable outcome, social serendipity, by definition, is a complete and successful experience that always results in benefits. It is worth pursuing to induce these experiences. Prior research suggested that whether individuals can experience serendipity depends on whether physical or digital environments possess key affordances for serendipity (Björneborn, 2017). As seen in the story example, contextual and social settings play a significant role in initiating the experience via various triggers that might have the capacity to induce serendipity. Figure 2.5 provides a synthesis of the triggers classification and the types of serendipity affordances, which were derived and interpreted from the works of de Melo (2018) and Björneborn (2017). Accordingly, triggers' capacity to enable access to diverse, dissimilar, and incomplete content refers to *Diversifiability* affordance. Coupled with an individual's ability to be open-minded, such an affordance can spark playfulness, curiosity, and interest, which are essential for noticing serendipitous cues. Next, if the triggers allow individuals' mobility and divergent exploration within the environment, they might possess the so-called *Traversability* affordance. This can motivate individuals to search for unsought, enable immersion, and stumble, essential activities to increase the likelihood of serendipitous occurrences. Finally, *Sensorability* affordance refers to triggers' ability to stimulate the senses and has a high degree of rich stimuli.

Coupled with individuals' sensitivity and attention, sensorability produces a sense of surprise and recognizes the serendipitous experience.

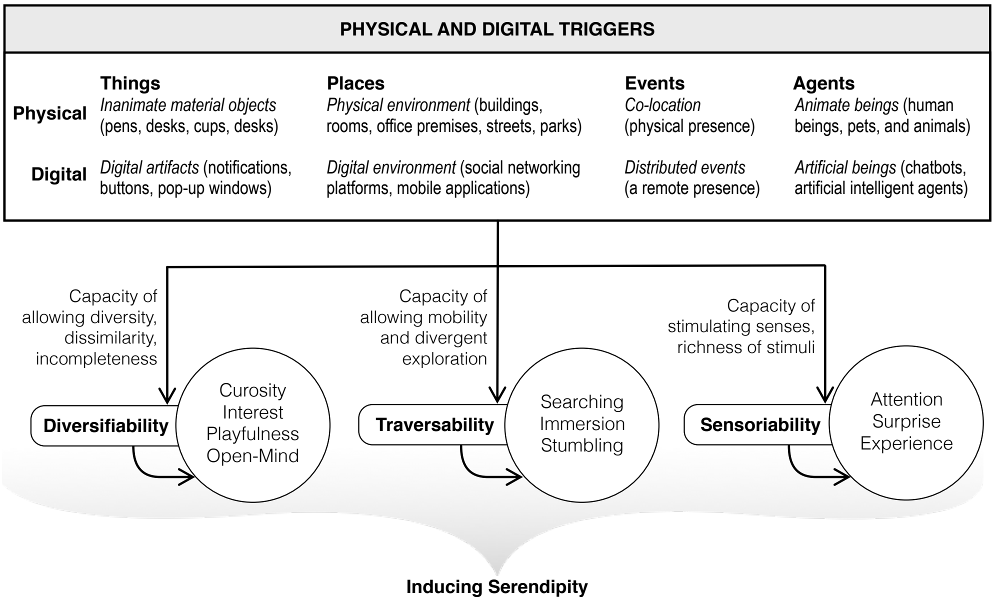


Figure 2.5 Triggers and key affordances for serendipity.

Well-known digital environments, such as Google and Twitter, have been recognized as having affordances for serendipity. Such services emphasize dynamism and diversity as critical qualities that open access to diverse information and different paths to encounter it (Lutz et al., 2013; McCay-Peet & Quan-Haase, 2016). A branch of research investigated how to design for serendipity, resulting in a debate about whether it is possible to develop technological features that enable serendipity (Lutz et al., 2013; Makri et al., 2014). RecSys research has chosen to replicate serendipitous qualities such as unexpectedness, novelty, and relevance (Adamopoulos & Tuzhilin, 2014; Iaquinta et al., 2008; Oku & Hattori, 2011; Y. C. Zhang et al., 2012) to amplify diversity and user satisfaction in content recommendations. Serendipity's potential value in matching people has been acknowledged but not empirically studied in RecSys (Pizzato et al., 2013). IR research took the opposite position, claiming that engineering serendipity is an oxymoron (André et al., 2009; McBirnie et al., 2016), but it is possible to facilitate the perception of serendipity through strategies that promote divergent exploration, seize opportunities, and relax personal boundaries (Makri et al., 2014). CSCW and HCI studies on social matching and networking have proposed various designs that support chance and spontaneous encounters and increase social awareness (Eagle & Pentland, 2005; Erickson & Kellogg, 2000; Jeffrey & McGrath, 2000), yet such approaches do not necessarily result in serendipity.

Although prior research acknowledges serendipity's importance in knowledge work, little attention has been paid to its social component, resulting in a lack of empirical understanding of related subjective experiences and the role of technology in unsought PSM. Various disciplines have examined the phenomenon and proposed designs for it, mainly within the context of information retrieval and content recommendations. This dissertation calls for a thorough exploration of social serendipity as a manifestation of successful PSM by presenting two studies on experiences of social serendipity in digital and physical environments.

2.2 Results: Key Factors and Antecedents of Social Serendipity

The findings indicate that knowledge workers were exposed to social serendipity across organizational boundaries in various contexts, including physical and digital environments, resulting in a broad spectrum of perceived benefits (see Table 2.1). The results also present a variety of contexts and stimuli that influence the social networking process. The ambiguity in distinguishing between serendipity and pure chance contributes to differences in the complexity of perception and essential factors within the reported processes of making unsought connections. Therefore, the findings conclude that social serendipity is a subjectively perceived, well-formed, coherent, and deliberately acknowledged experience that unfolds over time and embraces the joy of unexpected fortune. The primary difference between social serendipity and pure luck and chance is that it is a proactive, interactive, and reciprocal process that requires people to invest in and commit to reaching a beneficial outcome. Both studies illustrated that technology plays a limited role in social serendipity and concluded that the phenomenon is mainly associated with factors related to individual characteristics, such as personality and behavior.

While prior research on serendipity acknowledges only that personality characteristics play a role in experiencing serendipity (McCay-Peet & Quan-Haase, 2016), the present findings provide empirical evidence (see Table 2.2). The results of the first study revealed that people tend to experience social serendipity when there is no time pressure, and they have the internal capacity to socialize and remain open to

Table 2.1 Summary of experiential categories in perceived social serendipity and related examples from Publication I.

Experience category	Examples
Contextual trigger	• Public places (e.g., airports, clinics, train stations, hotels, bus stops) • Professional events (e.g., conferences, exhibitions, seminars) • ICT-mediated environments (e.g., phone and video calls, emails) • Work premises (e.g., kitchens, meeting rooms, offices) • Educational premises (five stories); social events (e.g., fairs)
Interaction triggers	• Direct communication (e.g., small talk, official meetings) • Repeated encounters • Unsought interactions • Receiving assistance or advice
Connection phase	• Revealed commonalities, e.g., job fields, professional and personal interests, hobbies, life goals • Revealed complementarities, e.g., professional skills and needs, different backgrounds and viewpoints
Follow-up phase	• Repeated interactions, e.g., multiple meetings or chats • Contact information exchanges
Valuable outcome	• Primary—professional, e.g., collaboration, vocational growth, business partnership • Secondary—social and personal, e.g., friendship, extended social networks, mentoring

opportunities. Following the Big Five Personality Traits (Gosling et al., 2003; Lang et al., 2011), the second study investigated the roles of openness to experience, neuroticism, agreeableness, conscientiousness, and extraversion in experiencing social serendipity (see Table 2.2). The findings statistically consolidated the importance of openness to experiences. This is an expected yet influential factor, because previous research conceptualized such openness as a "prepared mind," which is essential for recognizing and perceiving serendipity (e Cunha et al., 2010; McBirnie, 2008). Neuroticism is another crucial personality characteristic found to positively associate with social serendipity. Typically, neuroticism is seen as a negative trait that manifests as a tendency to experience negative emotions such as anxiety and moodiness (Barlow et al., 2014). However, there are positive aspects of neuroticism, such as an ability to mind-wander and shift attention to task-irrelevant thoughts (Robison et al., 2017). Such behavior would ensure more space for idle moments and associative states of mind necessary for noticing unsought yet valuable opportunities (Makri et al., 2014; McCay-Peet & Toms, 2015). The findings also illustrate that being conscientious (i.e., organized and task-oriented) is unfavorable to social serendipity—the less organized and self-disciplined the person, the higher the probability of encountering serendipitous contacts.

The first study demonstrates that most socially serendipitous experiences occurred in physical environments, and only seven of 37 cases featured ICT-mediated environments such as social network services, phone or video calls, or emails. The findings suggest that technology played an insignificant role by enabling chance encounters, while follow-up activities and value creation took place naturally without technological assistance. The second study statistically consolidated technology's minor role in social serendipity. Drawing from functional affordances theory (Hartson, 2003; McGrenere & Ho, 2000), the study addressed how intrusive features (enabling connectivity among users), usability features (the usefulness of an IT service), and dynamic features (the frequency of changes in the IT environment) associate with social serendipity. Such functional affordances refer to technology characteristics such as presenteeism, self-disclosure, recommendation quality, and pace of change that might help the user achieve socially serendipitous experiences. The study's primary assumption was that the quality of the recommendation, which aims to enrich the user experience in many ways, would also facilitate social serendipity. However, the findings reveal that, apart from presenteeism (the extent to which users believed that technology made them reachable and accessible), other characteristics did not affect social serendipity.

In summary, serendipity in social contexts, particularly in knowledge work, has been underexplored in empirical research. In contrast to prior research that focuses primarily on information serendipity, this dissertation unfolds the multifaceted na-

Table 2.2 Excerpt of Publication II findings illustrating antecedents of Information Serendipity (IS) and Social Serendipity (SS). N=473, unstandardized β values, CI= confidence intervals, * p <0.05, ** p <0.01, *** p <0.001

	IS β (sig.)	SS β (sig.)	IS (95% CI)	SS (95% CI)
(Constant)	.326	0.569	[-0.544, 1.195]	[-0.439, 1.577]
Background characteristics				
Gender	0.037	0.104	[-0.131, 0.205]	[-0.091, 0.298]
Age	-0.005	-0.008	[-0.014, 0.004]	[-0.018, 0.002]
Twitter use experience	**0.167*****	0.103	[0.047, 0.287]	[-0.036, 0.242]
Number of Followees	0.89	0.128	[-0.276, 0.016]	[-0.044, 0.299]
Number of Followers	**-0.130***	**-0.161***	[-0.059, 0.237]	[-0.330, 0.008]
Personality characteristics				
Openness to experience	**0.147*****	**0.158*****	[0.078, 0.234]	[0.067, 0.249]
Neuroticism	**0.080****	**0.075****	[0.015, 0.141]	[0.002, 0.148]
Agreeableness	-0.035	-0.018	[-0.144, 0.055]	[-0.134, 0.097]
Conscientiousness	**-0.074***	**-0.113****	[-0.151, 0.012]	[-0.207, -0.018]
Extroversion	0.001	0.025	[-0.065, 0.062]	[-0.049, 0.098]
Twitter charcteristics				
Presenteeism	**0.593*****	**0.508*****	[0.472, 0.696]	[0.378, 0.638]
Self-disclosure	-0.017	0.046	[-0.115, 0.080]	[-0.067, 0.159]
Recommendation quality	0.067	0.064	[-0.018, 0.162]	[-0.040, 0.169]
Pace of change	-0.031	-0.004	[-0.125, 0.077]	[-0.121, 0.113]
Types of Twitter use				
Professional use	**0.181*****	**0.144****	[0.072, 0.286]	[0.020, 0.268]
Receiving	**0.191****	-0.038	[0.000, 0.318]	[-0.222, 0.146]
Broadcasting	0.060	0.007	[-0.084, 0.235]	[-0.179, 0.192]
Interacting	**-0.252*****	0.044	[-0.385, -0.108]	[-0.116, 0.205]
R squared	0.37	0.27		

ture of social serendipity. The analysis revealed qualitative and quantitative insights into various aspects, characteristics, and antecedents of social serendipity. The complex nature of serendipitous experiences suggests the need to identify new mechanisms to promote serendipity in professional contexts.

3 DESIGNING DIVERSITY-ENHANCING STRATEGIES

This chapter synthesizes the findings of Publications III, IV, and V and aims to understand social data sources and diversity-enhancing strategies for people recommendations (see Figure 3.1). Publication III is an exploratory literature review that advances the understanding of implicit social data types that can be used to profile and model users. Understanding the sources of BSD, in turn, can guide analytical procedures for designing diversity-enhancing recommendation strategies, which are presented in experiment-based studies in Publications IV and V. I provide an overview of people recommender systems and existing recommendation approaches in the following. Next, the description of proposed diversity-enhancing strategies is presented.

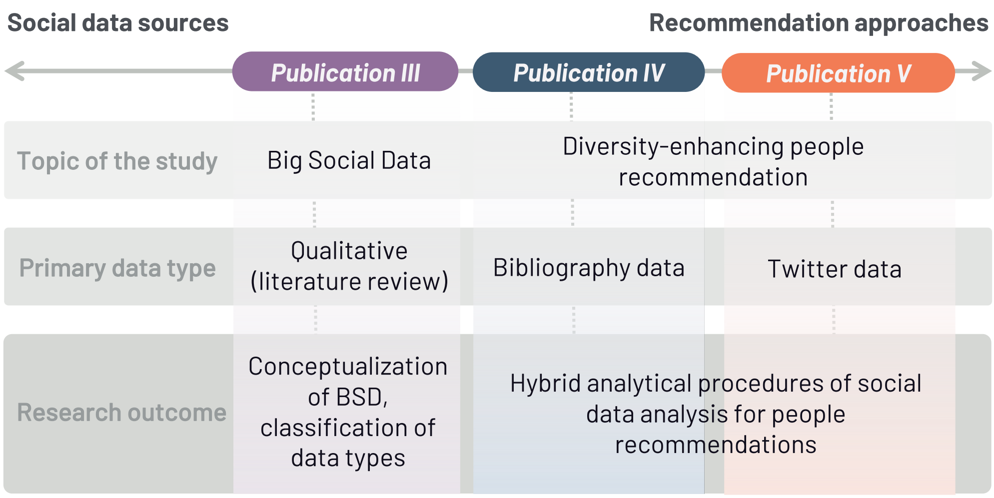


Figure 3.1 Overview of Publications III, IV, V highlighting differences among the studies in relation to research focus, topic, data types, and outcomes.

3.1 Background: Recommender Systems and Recommendation Approaches

3.1.1 People Recommender Systems

With the emergence of social media services and pioneering research on social matching by Terveen and McDonald (2005), people-to-people recommendations became the distinct domain of the social recommenders (Guy, 2015). In his recent research, Guy (2018) proposed a relationship-based taxonomy of people recommendation systems that distinguished social matching approaches based on:

- *Familiarity*—recommending known actors like friends and colleagues (Guy et al., 2009; L. Zhang et al., 2011);
- *Interest*—establishing unidirectional relationships with familiar but not necessarily known actors (e.g., following a celebrity or finding an expert) (Hannon et al., 2010; Yogev et al., 2015);
- *Similarity*—introducing strangers with similar interests to grow the network (Guy et al., 2011; Lopes et al., 2010).

This dissertation mainly subscribes to the third type of matching yet is not limited to a singular similarity or familiarity level. Instead, this work approaches different kinds of recommendations from the perspective of recommendation strategies and their features (see Figure 3.2).

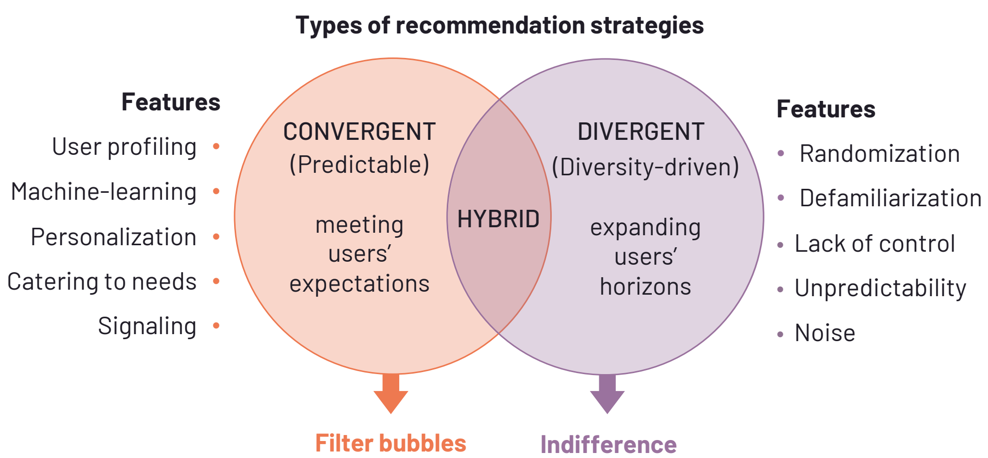


Figure 3.2 Types of recommendation approaches and corresponding features.

Inspired by the classification of convergent and divergent systems suggested by de Melo (2018), I distinguish between convergent and divergent recommendation strategies. Convergent systems cater to users' needs, tastes, and interests, while divergent systems tend to neglect previously mentioned and provide random suggestions. Most of the social networking tools discussed in Section 2.1.2 represent examples that utilize convergent strategies for professional matchmaking. Many expert-finding systems use convergent strategies. They provide recommendations based on targeted search criteria, prioritizing those that fit the query (Al-Taie et al., 2018; Lin et al., 2009). Convergent recommendations focus on meeting and predicting user expectations through personalization. Such strategies strongly depend on user profiling and modeling and analyze users' explicit input data and implicit interaction- or behavior-based signals. While these strategies aim to achieve maximal relevance of recommendations, they are highly likely to trap users in filter bubbles (Pariser, 2011) and provide similar suggestions (in relation to one another).

In contrast, divergent strategies promote diversity exposure (Helberger et al., 2018), prioritizing the breadth of the recommendation pool over the relevancy. Typically, such a strategy is achieved through randomization and defamiliarization techniques (Bardzell, 2013) by limiting the user's role or agency in affecting the filtering mechanisms. The opportunistic social matching application "Next2You," which recommends nearby strangers and builds on progressive disclosure (Paasovaara et al., 2016), applies a divergent strategy for social networking. Users have no control over the recommendations they receive and passively collect interesting facts about encounters in proximity. Thus, divergent strategies are characterized by high unpredictability, resulting in a high rate of irrelevant suggestions and subjective indifference toward them.

I consider hybrid recommendation strategies to be a diversity-enhancing optimum. They utilize diversity exposure, also taking into account the relevancy dimension. The following provides examples of such strategies and addresses related recommendation approaches.

3.1.2 Existing Recommendation Approaches

All recommendations essentially start from the users' input data. Based on the literature review (Publication III), I identified three primary data types that can be collected and further used for profiling, modeling, and filtering mechanisms. The first type is *digital self-representation data*, which refers to explicit information that users share to socialize and express themselves online. This data type includes login data, typically used as an identifier to locate users in the global dataset and personalize recommendations. Location data can also be used as a filtering dimension to, for instance, recommend people who are in proximity. This data type also includes self-published content, such as posts, tweets, and various media, which is often the target data to analyze interests, expertise, and knowledge to produce recommendations. The next type is *technology-mediated communication data*, which relates to traceable social interactions between users, such as mentions, comments, and likes, which could be used to identify interaction patterns. The last type is *digital relationships data*, widely used in recommendations to filter out existing connections and identify new ones. It could be explicit—visible in the list of the user's connections (e.g., friendship, followership), and implicit, derived from interaction patterns and community network structures using social network analysis methods.

Following such a data classification, traditional approaches to people recommendations can be categorized into content-based (content analysis) and network-based (social connection analysis). Typical content-based approaches compare profiles to reveal topical similarities or shared interests (Guy et al., 2011; Lopes et al., 2010; Van Le et al., 2014). They often use text-based analytical procedures such as TF-IDF (Beel et al., 2013) and LDA (Blei et al., 2003) and weight similarity according to the cosine distance (Li & Han, 2013). A specific case study on Twitter illustrated how to match people based on the similarity of popularity and activity by comparing the ratios of users' followees and followers and the number of tweets (Garcia-Gavilanes & Amatriain, 2010). Network-based approaches analyze social links between people and predict potential new connections. A well-known network-based approach is based on the triadic closure principle (friend-of-friend connection) (Carullo et al., 2015), which Facebook and LinkedIn use. There is also an approach based on the weak ties theory (Granovetter, 1983), which aims to enhance networks' structural diversity by bridging existing communities with strong ties (Sanz-Cruzado & Castells, 2018).

Since traditional approaches have been criticized for supporting homophily bias (Kossinets & Watts, 2009) and causing social polarization (Fuchs, 2017), RecSys research has started to promote and employ hybrid approaches as a diversification strategy. Primarily, hybrid approaches in people recommendations can be classified into three diversity-driven categories: diversity of features, analytical procedures, and data.

Diversity of features is the most conventional approach and focuses on obtaining multiple features from the user: explicit or implicit characteristics such as interests, social networks, and affiliations, among others. A representative example is a recommendation system by Tsai and Brusilovsky (2019), which builds on the beyond-accuracy objective, addressing the dynamic needs of social matching among scholars. The authors proposed extracting four features:

- *Academic feature*—the degree of topical publication similarity;
- *Social feature*—social circle similarity among academics, combining common neighbor networks and co-authorship;
- *Distance feature*—geographical distance among scholars based on affiliation;
- *Interest feature*—identification of co-bookmarked articles and connected authors.

Diversification of the recommendations pool is achieved by combining pairs of features, resulting in intersections of their relevance. In addition to the diversity of features, the authors designed a diversity-enhancing visual interface, "Scatter Viz," that facilitates exploratory investigation of potential matches. Another example, Amal et al. (2019) proposed building a relational social graph that allows feature extraction, including interests, affiliation, residence, and awards, to match scholars. There is also a case study on extracting contextual features, such as mobility and activity, for followee recommendations on Twitter (Q. Yuan et al., 2015).

Diversity of analytical procedures refers to utilizing hybrid analysis techniques to filter the recommendation pool. A typical approach is to combine content and social network analyses to increase the relevance of recommendations (Kong et al., 2016). This can include Brusilovsky et al. (2017) demonstrated synergy of content-, tag- and social-based methods and analysis of publications' metadata, tags, and co-authorship networks. Another example complements the identification of content similarity with sentiment analysis, which allows recommending followees on Twitter with either similar (Akiyama et al., 2017) or different (Gurini et al., 2013) emotional responses to a common topic of interest.

Diversity of data combines various data sources for profiling, modeling, and filtering. The domain of matching scholars at conferences was the primary contributor to this category of hybrid approaches. For instance, the "Conference Navigator" system (Tsai & Brusilovsky, 2016) extends user modeling by fetching Google Scholar, ResearchGate, Twitter, and Wikipedia data using the Google Custom Search API and further employing content and network similarity measurements. Arens-Volland and Naudet (2016) implemented a mobile application prototype, "Adaptive Conference Companion," enriching users' profiles by social network mining, including analyses of LinkedIn, ResearchGate, MyScienceWork, Google Scholar, and DBLP data. Explicit (users' input) and implicit (results from social network mining) information allowed the extraction of various features, such as skills, topics of interest, affiliations, residence, disciplines of interest, event objectives, and others. Combining data sources is a powerful approach for enriching the diversity of recommendations, but it is costly and limited by complications due to the recent General Data Protection Regulation (GDPR).

3.2 Results: Proposed Diversity-enhancing Strategies

Considering practicalities and available resources for the experiment-based research, this dissertation contributes to the diversity of analytical procedures by combining content- and network-based approaches to diversify the recommendation pool. As publication data and Twitter data were available for research without restrictions, they were selected as primary data sources for producing people recommendations.

Publication IV proposes a recommendation strategy based on the analysis of the similarity-difference continuum (see Figure 3.3). The DBLP dataset of Computer Science publication records was used to produce recommendations for scholars. The topics of interest were modeled with TF-IDF, which allows the formation of feature vectors needed to measure the similarity between actors (i.e., the calculation of cosine distances). The novel aspect of the strategy is to apply the OTSU filter (Otsu, 1979) to identify similarity thresholds and filter the pool of recommendations based on three degrees of similarity—high, moderate, and low. This beyond-maximal-similarity approach is especially relevant in the scholarly domain due to the diverse needs of collaboration, which can require complementary or varied areas of expertise.

Publication V demonstrates how to diversify the recommendation pool by identifying topology-based structural network positions (see Figure 3.3) using Twitter data

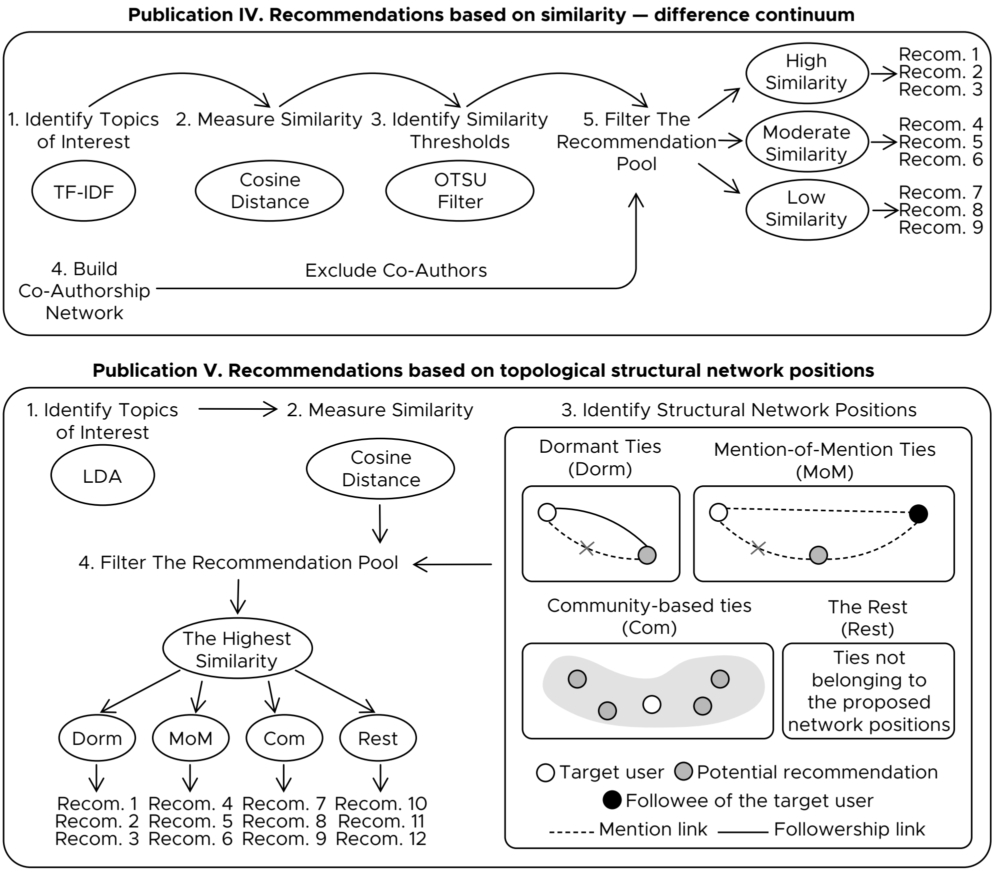


Figure 3.3 Diversity-enhancing strategies based on similarity-difference continuum and topological structural network positions proposed in Publications IV and V.

to recommend scholars and stakeholders at local universities in Tampere, Finland. The user modeling was performed with an unsupervised topic modeling algorithm LDA. As the primary goal was to assess the structural network positions as filtering dimensions for followee recommendations, the cosine distances were calculated to retrieve the highest similarity, so it would not affect the evaluations. The novelty of the proposed strategy lies in combining mention-based and followership-based networks to identify different types of structural network positions:

- *Dormant ties*—followership with no explicit interactions;
- *Mention-of-mention*—a friend-of-friend connection in a mention network;
- *Community-based*—users belong to a shared community with no explicit followership or interactions.

Producing such recommendations may help decrease social polarization in online networks and expand social circles by exposing users to connections they would otherwise miss in contemporary systems that employ convergent strategies.

4 EVALUATING PEOPLE RECOMMENDER SYSTEMS

This chapter summarizes findings on subjective perception regarding recommendations produced with proposed diversity-enhancing strategies from Publications IV and V. Publication IV evaluated scholars' perceptions of relevance about potential collaborators with three levels of similarity viewed from the perspective of computed content distance. Publication V aimed to assess perceptions of followee recommendations on Twitter from the perspective of different network structural positions. Although how the recommendations were produced and the empirical contexts were distinct, both publications followed a user-centered approach, focusing more on the user experience than traditional accuracy measures to identify recommendation efficiency. Therefore, the evaluation was based on subjective attitudes and opinions collected through custom measures for variables and included perceived relevance, familiarity, similarity, and willingness to interact or follow up on recommendations. In the following, I first address the relevant literature on recommender system evaluation and then summarize the results.

4.1 Background: Evaluation Approaches and User Experience in Recommender Systems

4.1.1 Systems-Centric vs. User-Centric Evaluation Approaches

The goals and purposes of recommender systems vary significantly, which leads to challenges for systematic and unified approaches in evaluating them (Herlocker et al., 2004). For instance, some recommenders are designed to test new algorithms, and others focus on improving the user experience while interacting with recommendations. Evaluation techniques for such diverse objectives need to be tailored accordingly. Prior research suggests classifying evaluation approaches into *system-centric* (i.e., evaluating algorithmically) and *user-centric*(evaluating with user-centered studies or experiments) evaluation (Cañamares et al., 2020; Cremonesi et al., 2013).

Usually, system-centric evaluation uses a prebuilt set of user opinions on recommendation items as a ground truth (Cremonesi et al., 2013). Users do not interact with the system, and the evaluation aims to measure the recommendations' accuracy by comparing the ground truth datasets against the recommender system's predictions. The user-centric evaluation aims to observe the recommender system's use by actual people. Measurements can be collected through interviews, surveys, and controlled experiments. The system-centric evaluation approach has historically been more established, yet recent findings illustrate the increasing importance of user interaction in evaluating recommender systems (Cremonesi et al., 2013). System-centric approaches have been shown to overlook essential factors that affect the adoption and perceived usefulness of recommendations (Carraro & Bridge, 2020; Knijnenburg et al., 2012). Therefore, many researchers turned their attention to the role of the user experience in recommender systems (Champiri et al., 2019; Knijnenburg et al., 2012; Konstan & Riedl, 2012).

The system-centric approach is often chosen over the user-centric approach, as it is more cost- and time-efficient (Cremonesi et al., 2013). However, only user-centric evaluation can reveal insights on the perceived relevance of recommendations and the overall user experience, which many argue is more important than the pure accuracy measure (Buder & Schwind, 2012; Knijnenburg et al., 2012; Pu et al., 2012). Recommendations can be accurate but not always useful in supporting users' decision-making. We might encounter such flaws every day in online recommenders. For example, if I buy a black towel, I will see only black towels in my recommender feed or targeted ads. It is accurate that I like black towels, but I already bought one, so these recommendations are no longer relevant. However, this is exactly how accuracy metrics (Shah et al., 2017) such as prediction accuracy, decision support accuracy, and rank-recommendation metrics test recommenders. Such evaluations can lead to homogenization (Konstan & Riedl, 2012), where popular and previously rated recommendations are prioritized over less popular items or those with limited historical data (Cremonesi et al., 2011).

That said, system-centric evaluation approaches cannot evaluate all types of recommender systems, as accuracy alone does not ensure a positive user experience and relevant recommendations (Cremonesi et al., 2013; McNee et al., 2006). Previous research illustrates a lack of correlation between algorithmic accuracy and an improved user experience: Statistical accuracy metrics inconsistently predict recommendations' perceived quality (Cremonesi et al., 2011). Research has proposed that, even at the expense of accuracy and precision, recommenders should prioritize user needs (Bobadilla et al., 2013).

Prior literature suggested that recommender systems consist of three primary parts (Xiao & Benbasat, 2007):

1. *Input*—elicits users' preferences;
2. *Process*—generates recommendations;
3. *Output*—delivers recommendations.

System-centric metrics primarily focus on the process, while input and output are scarcely addressed (Xiao & Benbasat, 2007). However, they are crucial to the user experience (Pu et al., 2012). Eliciting user preferences is the first phase, where the user interacts with recommenders via searching, browsing, or purchasing. Next, the delivery phase is crucial to the user's first subjective perceptions and impressions. Finally, users can often provide feedback on recommendations by selecting or rating items, which reveal user preferences. Knijnenburg et al. (2012) also emphasized that the user experience is not only affected by the system. Contextual and personal characteristics significantly influence subjective perceptions. Therefore, evaluating how recommendations are perceived is insufficient, and it is important to assess whether user needs were met to understand the user experience comprehensively (Herlocker et al., 2004; McNee et al., 2006).

The most significant advantage to user-centric approaches is that the gathered insights help us understand what constitutes good or bad recommendations and why they are perceived as such (Konstan & Riedl, 2012). However, user-centric approaches have been used less than system-centric approaches. According to Konstan and Riedl (2012), unlike system-centric approaches, user-centric approaches demand resources: A system or prototype must be developed with algorithms and the user interface to carry out field studies or experiments with actual users and measure their behavior. In addition, the user experience in recommender systems is not well-defined because of insufficient established evaluation methods and metrics (Knijnenburg et al., 2012) and the complexity of operationalizing variables to measure or control (Cremonesi et al., 2011).

In summary, user-centric evaluation approaches can be challenging to implement but worth pursuing, given the value of impact. In the following, I focus on the user experience in recommender systems, leaving system-centric approaches behind, as all the publications in the dissertation utilized the user-centric evaluation approach.

4.1.2 What Makes a Recommender System "Good"? The User Experience in Recommender Systems

Previous research concluded that recommender systems must be evaluated in terms of the user experience because they are designed to meet user needs (Herlocker et al., 2004; McNee et al., 2006). This can be accomplished by collecting and analyzing users' subjective perceptions. Pu et al. (2012) provided an overview of metrics beyond accuracy, emphasizing the importance of subjective perceptions in evaluating recommender systems. The research addressed a variety of measures for subjective perceptions of the system's qualities, such as perceived accuracy (Pu et al., 2011), familiarity (Sinha & Swearingen, 2002), novelty and diversity (Cremonesi et al., 2013; T. T. Nguyen et al., 2014), and serendipity (Kotkov et al., 2016).

It is important to define the user experience in relation to recommender systems. Konstan and Riedl (2012) define it as *"the delivery of the recommendations to the user and the interaction of the user with those recommendations."* Knijnenburg et al. (2012) proposed that the user experience is *"the user's evaluation of the system (perceived system effectiveness and fun), system usage (usage effort and choice difficulty), and outcome of system usage (satisfaction with the chosen items)."* Neither of these definitions addresses the elicitation of user preferences. They emphasize experiences related to the outcome (presenting recommendations).

Leino (2014) considers the ISO definition of user experience (for Standardization, 2010) as a starting point, referencing the concept of anticipation, which has not been stressed in the RecSys field: *"A person's perceptions and responses that result from the use and/or anticipated use of a recommender system, where these perceptions and responses can be characterized in terms of effectiveness, efficiency, and satisfaction in achieving specified goals, and also in terms of an engagement with the recommender system."* This definition describes the interaction holistically, incorporating the emotional, aesthetic, and persuasive aspects of the user experience.

Established dimensions of user evaluations such as explicit versus implicit data gathering, laboratory versus field studies, outcome versus process, and short-term versus long-term studies allow researchers to study the user experience of recommender systems and their related aspects empirically (Herlocker et al., 2004):

- Explicit versus implicit data collection—asking users about their experience via interviews and questionnaires or observing their interactions and behavior. Research suggests that explicit and implicit methods should be combined to enrich the gathered insights (Knijnenburg et al., 2012; Knijnenburg et al., 2011). Behavioral data alone cannot always explain subjective experiences and can be biased by the observer. Explicitly asking the user might help interpret users' interactions (Kobsa, 2007).
- Laboratory versus field studies—laboratory studies are useful for testing hypotheses and controlling specific variables, while field studies bring contextuality to the experience. Often it is hard to find participants for laboratory studies who have an actual need and motivation to use a system, which affects the decision-making experience (Sillence & Briggs, 2007).
- Outcome versus process—refers to operationalizing metrics to focus on the successful outcome or the process of the experience. In a recommender system, this could relate to measuring whether the user received relevant recommendations and the decision-making process in choosing them.
- Short-term versus long-term evaluations—the evaluation should consider carefully whether research questions can be answered with short- or long-term studies. For example, the system's adoption can only be ascertained with long-term studies.

The following section explains how the user-centric approach was utilized in Publications IV and V. I also explicate the operationalization of subjective measures of familiarity, relevance, and follow-up activities specifically tailored to people recommenders.

4.2 Results: Evaluating People Recommendations Beyond Accuracy

The previous section addressed the interest in user-centric evaluation approaches in the RecSys field. However, there are still no unified, specific measures to evaluate people recommenders. When recommending people, the concept of recommendation quality is widely interpreted. Therefore, conducting experiments on proposed diversity-enhancing recommendation strategies for PSM required operationalizing

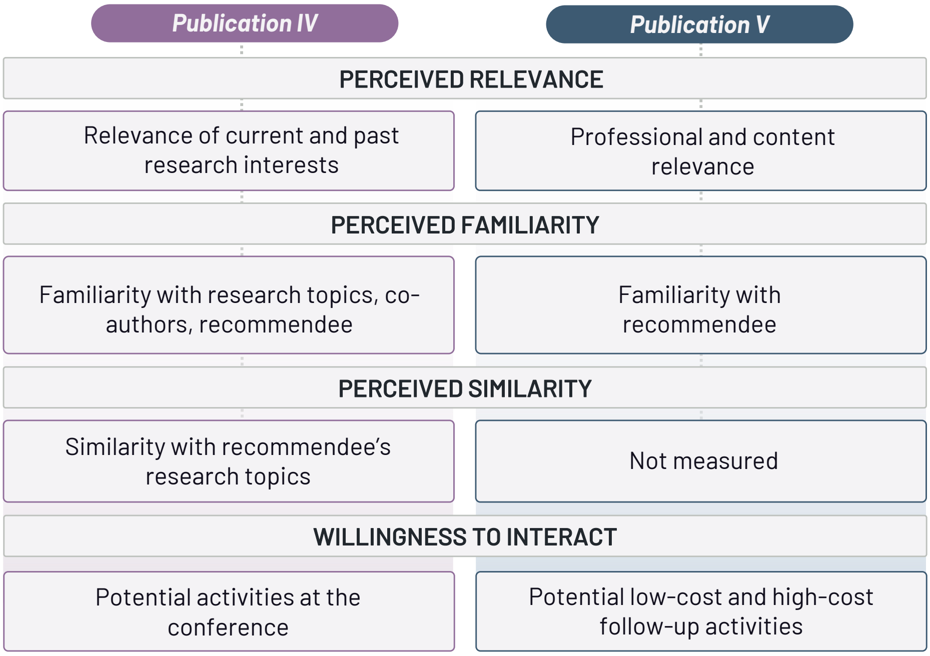


Figure 4.1 Proposed measures used in Publications IV and V and their comparison.

such subjective measures. Figure 4.1 presents four primary constructs, such as perceived relevance, familiarity, similarity, and willingness to interact, and compares how they differ in two experiments.

Table 4.1 demonstrates examples of statements used to measure the constructs. Such measures uncover experiential aspects and quantitatively assess how individuals perceive people recommendations. Specifically, subjective measures consolidated objective findings, indicating the distinct nature of the recommendations produced with proposed diversity-enhancing recommendation strategies (see the summary of findings in Figure 4.2). The fact that respondents identified relevant connections from different content and network structure similarity-difference continuums suggests that the proposed recommendation strategies are meaningful approaches for diversifying people recommendations.

To complement reductionist measures, qualitative findings collected through semi-structured interviews (Publication IV) and open-ended questions (Publication V) reveal more complex and nuanced aspects that should be addressed in the design and evaluation of people recommendations for PSM. The qualitative feedback about recommendations of different similarity-diversity levels aligned with the quantitative re-

Table 4.1 Examples of measure statements used in studies.

Construct	Publication IV	Publication V
Perceived relevance	I consider this person relevant to me from the perspective of my current research interests	This person shares information and content that I find useful; I find this person interesting for my professional activities
Perceived Familiarity	I know by name some of the co-authors of this person	Scale: (1) Very unfamiliar - (5) Very familiar
Perceived Similarity	Scale: (1) Very different – (7) Very similar	Not measured. Objective measures of similarity were used instead
Willingness to interact	I consider this person relevant to me for the following activities in the context of a conference: asking for advice; exploring joint research interests; organizing a research visit, etc.	Low-cost interaction: I intend to start following this person on Twitter; High-cost interaction: I intend to contact this person for a face-to-face meeting

sults. An overview of participants' comments is presented in Figure 4.3. Although the participants were unaware of different recommendation groups, in their feedback, they distinguished between different degrees of perceived relevance by using phrases like *'very/most relevant,'* *'somewhat relevant/not an exact match,'* and *'irrelevant/totally irrelevant.'* Qualitative findings also reveal that homophily bias is evident in intuitive assessments of relevance and willingness to interact, and there is a mismatch between people's intuitive choices and deliberate intentions in decision-making when choosing relevant potential connections. Some participants mentioned that their first impression of recommended people had changed when they started to check the recommendation profiles thoroughly:

> *That was a good idea that you gave me to check all recommendations first because initial reaction was different, but when I started thinking and realized that my first impression maybe was not correct. When you start thinking about recommendations' relevance further, there might be some changes. So some of them are not that irrelevant as I thought at first.* (Excerpt of participant's feedback, Publication IV)

> *An interesting personality. I was able to see that we have something in common only after a closer look at the profile. Good Tweets and Retweets. Also,*

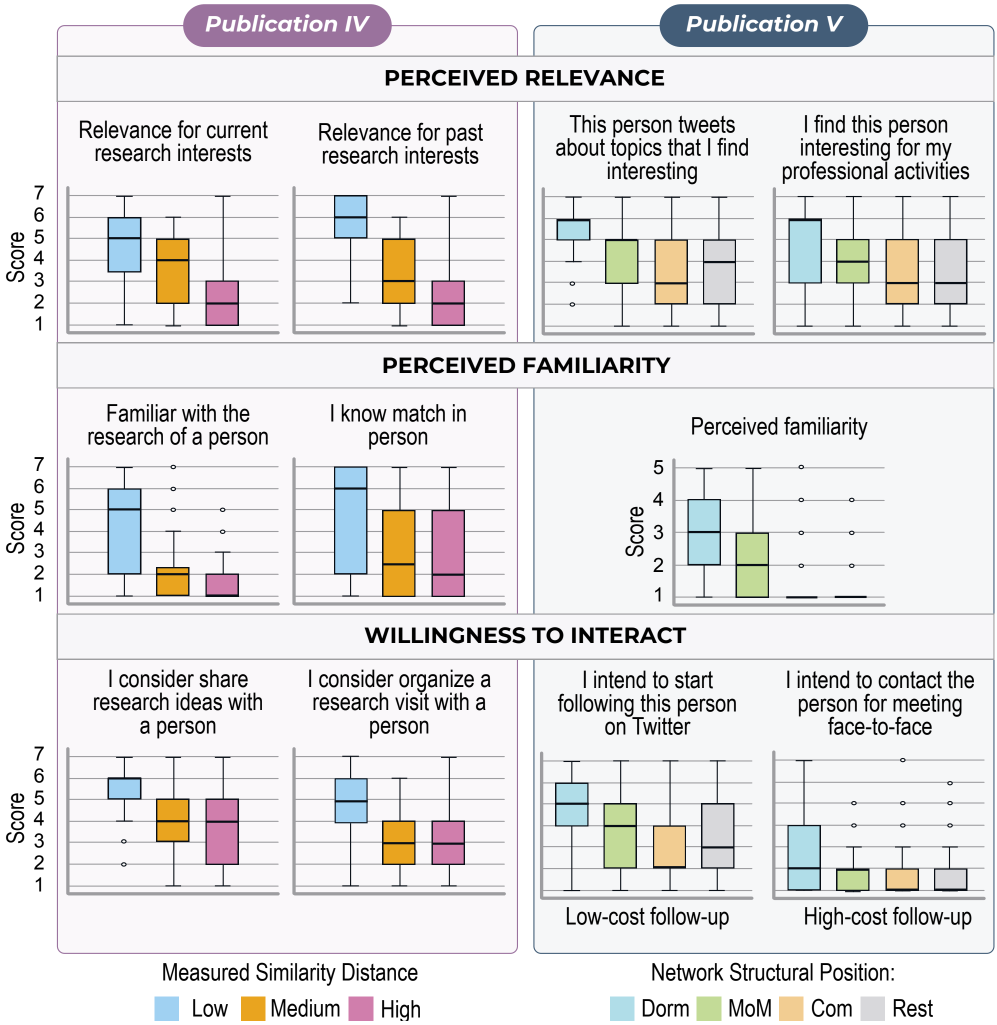


Figure 4.2 Excerpt of quantitative findings from Publications IV & V illustrating the distribution of participants' scores regarding perceived relevance, familiarity, and willingness to interact.

> *the fact that we have 88 shared followers creates trust.* (Excerpt of participant's feedback, Publication V)

Regarding needs and expectations in PSM, the results demonstrate that the optimal area on the similarity-difference continuum highly depends on the type and context of the envisioned collaboration. In short-term and low-cost professional interactions, factors like personal compatibility and similarity of attitudes receive less emphasis compared to long-term and high-cost collaborations. Additionally, the

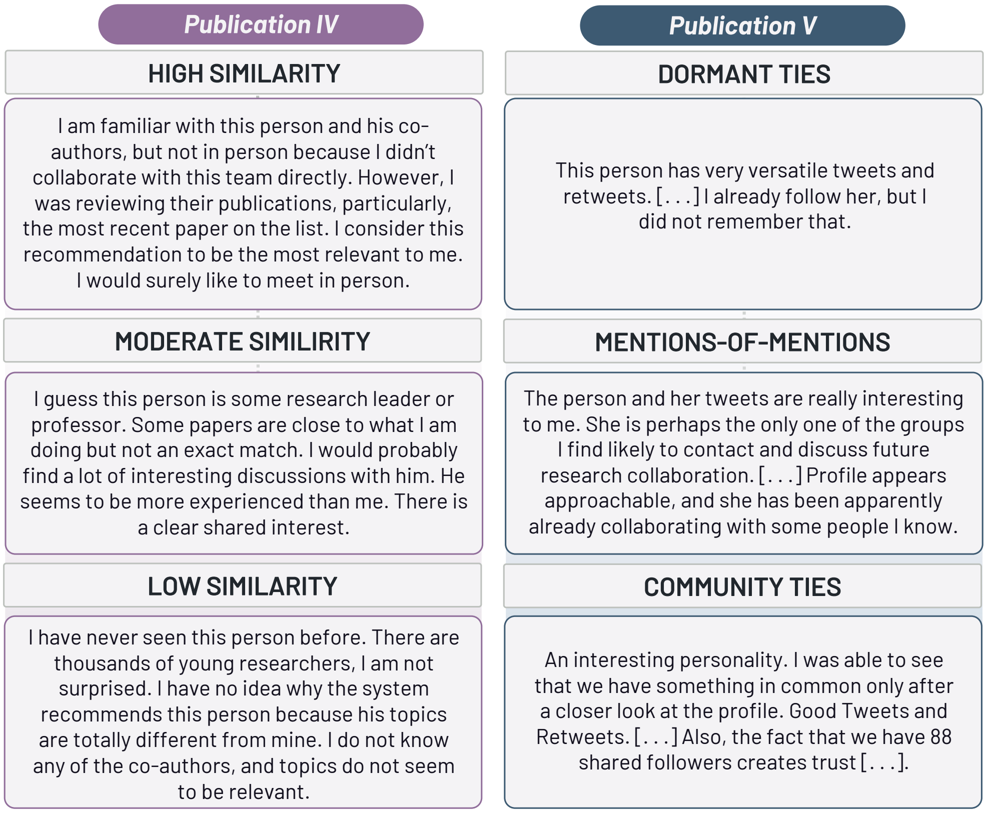


Figure 4.3 Excerpt of qualitative findings from Publications IV & V illustrating participants' opinion regarding different types of recommendations

nature of the collaboration task can impact the perceived relevance of a potential connection, particularly regarding the complementarity of professional roles, skills, and knowledge.

The findings of the user-centered evaluation allowed us to define several aspects that encompass the relevance criteria in PSM. First, *similarity* is essential in PSM in background, attitudes, values, beliefs, goals, and professional aims, as it promotes cohesion and mitigates individual dissimilarities in collaborative work. Second, *complementarity* in professional roles, skills, knowledge, and social capital identifies intersections that facilitate beneficial collaboration. Third, *compatibility* in direct cooperation involving mental, social, moral, and emotional closeness is needed to establish trust and valued cooperation. Finally, *approachability/logistics* considered physical and organizational proximity and social network structure, emphasizing the importance of smooth communication and interaction.

In summary, while the experiments indicated consistent and significant differences in subjective perceptions of the proposed diversity-enhancing strategies, the results imply that evaluating the relevance of people recommendations is complex and multifaceted. We operationalized the concept of perceived relevance, familiarity, similarity, and willingness to interact within the context of evaluating prospective PSM. By showing how these variables match with system-based objective measures, we revealed the asymmetry of intuition-based evaluation and deliberately considered intentions. Despite the inherent bias in selection, participants could identify relevant others at all levels of similarity and structural network positions.

5 DISCUSSION AND CONCLUSIONS

In conclusion, the dissertation focused on developing and evaluating PSM recommendation strategies that emphasize diversity and serendipity. This multidisciplinary and exploratory research consisted of five publications investigating the role of serendipity in making valuable connections, exploring Big Social Data as a means of user modeling and social recommendations, and proposing recommendation strategies to enhance diversity and relevance in PSM. The findings highlighted the importance of proactive and reciprocal sense-making in assessing the relevance of new connections and shed light on the qualities that facilitate serendipitous encounters in professional settings. The recommendation mechanisms proposed in this work offer a similarity-difference continuum of user interests and identify diverse network structural positions, expanding the range of potential matches. User-centered evaluations revealed the influence of human biases, such as homophily and familiarity, in decision-making processes when choosing relevant recommendations. Thus, insights reveal the mismatch between intuitive choices and deliberate rationalizations regarding recommendation relevance. We advocate for design-oriented research to discover recommendation strategies that can improve the likelihood of long-term beneficial PSM. In the following, I first revisit the dissertation's research questions and objectives. Then, I outline design considerations for diversity-enhancing strategies to be used in PSM tools and applications. Finally, I address limitations and future work.

5.1 Revisiting Research Questions and Objectives

The outcomes of the dissertation demonstrated that PSM involves proactive and reciprocal decision-making regarding the relevance of making new connections, which lack, yet would benefit from, further technological facilitation (Publication I). The dissertation also provided a detailed account of the various experiential, contextual, and personality qualities that favor serendipitous encounters related to work (Publications I and II). The findings also illustrate how BSD could be used for more socially acceptable purposes, that is, to identify and facilitate new social connections between people (Publication III). The recommendation mechanisms proposed in this dissertation modeled users on the similarity-difference continuum and identified various network positions between any given users, thus diversifying the pool of prospective matches (Publications IV and V). The results indicate the prevalence of human biases in evaluating recommendations: homophily and familiarity have a substantial effect when people decide with whom to connect (Publications IV and V). Furthermore, there is an apparent mismatch between intuition-based and deliberate rationalizations regarding the relevance of matches. Although the needs of professional networking call for heterogeneity and complementarity, intuition-driven choices favor those closest and most similar (Publication IV).

This research accentuates the need for technology-facilitated PSM and highlights the significance of diversity and serendipity in professional networking. To answer the first research question, social serendipity can occur in various settings and involves a degree of spontaneity and unpredictability, enhancing creativity and innovation in knowledge work. The results of both studies contribute to research on PSM within the CSCW and HCI fields by advancing the understanding of the impact of personality, behavior, and technology on social serendipity. Unfortunately, the results illustrate the minor role of technology in supporting such experiences, which calls for revisiting computational approaches and mechanisms to promote serendipitous social encounters more effectively. I believe that understanding key characteristics and antecedents of social serendipity informs the design space for PSM tools and recommender system mechanisms that facilitate chance encounters and promote growth and maintain professional relationships with higher rates of valuable outcomes. Thus, the present research calls for inducing serendipity using technology in new ways to assist knowledge workers in building meaningful collaborations that otherwise would be missed in conventional networking channels.

For the second research question, the findings contribute to interdisciplinary research on social and people recommender systems by proposing non-conventional perspectives to diversify the pool of people recommendations. Objective observations of proposed strategies extend the discussion of analytical mechanisms and algorithmic measurements beyond content-based similarity and triadic closure. The experiments demonstrated that diversity-enhancing objectives could be achieved using alternative strategies to well-established and widely used concepts in RecSys, such as similarity and social network structures. While it might be assumed that recommending people would require large and diverse data, the results illustrate that the recommender engine could function even with scarce data if the filtering mechanisms are properly designed. In particular, the findings reveal that participants could locate relevant recommendations of varied similarity levels and structural network positions. This provides a preliminary indication of the proposed strategies' potential for diversifying recommendations without losing relevancy.

Finally, in response to the third research question, subjective measures were operationalized to evaluate proposed diversity-enhancing recommendation strategies with the user-centric approach. The findings contribute to understanding subjective perceptions of relevance, complementarity, and diversity in enriching one's professional networks. The gathered insights inform how to strategize and optimize diversity and similarity within the design of people recommenders. However, measuring potential follow-up activities beyond the intention to follow a recommended person was challenging, as respondents showed little interest in high-cost activities like face-to-face meetings. This raises questions about using such measures as indicators of recommendation quality, particularly in controlled experiments. Furthermore, measuring the relevance of people recommendations is inherently challenging, as the long-term value of more diverse social networks may not be reflected in immediate impressions.

5.2 Design Considerations: Towards Diversity-enhancing Strategies

The following highlights design considerations and research questions, particularly for new-generation PSM systems and people recommender systems, which go beyond merely delivering contact suggestions. Such systems represent a more proactive technology paradigm, which utilizes an algorithmic logic for modeling and analyzing social relevance and actively assists users' decisions by making complex inferences.

I subscribe to the conceptualization of serendipitous systems (de Melo, 2018), which aim to facilitate all phases of a serendipitous experience from initiation to obtaining valuable outcomes, thus ensuring discoveries, unexpectedness, and value. Accordingly, I propose the design considerations around four primary computational process-based service qualities focusing on aspects relevant to the HCI (See Figure 5.1):

- Introducing opportunities for professional networking;
- Supporting the assessment of social recommendations' instrumentality by delivering cues on their inferred relevance through enhanced profiling and presentation;
- Encouraging direct social interactions for value creation by applying encouragement and persuasion techniques, thus assisting follow-up activities and trust-building;
- Maintaining social awareness of established connections to reach relationship maturity and integrity and provide cues on potential valuable outcomes.

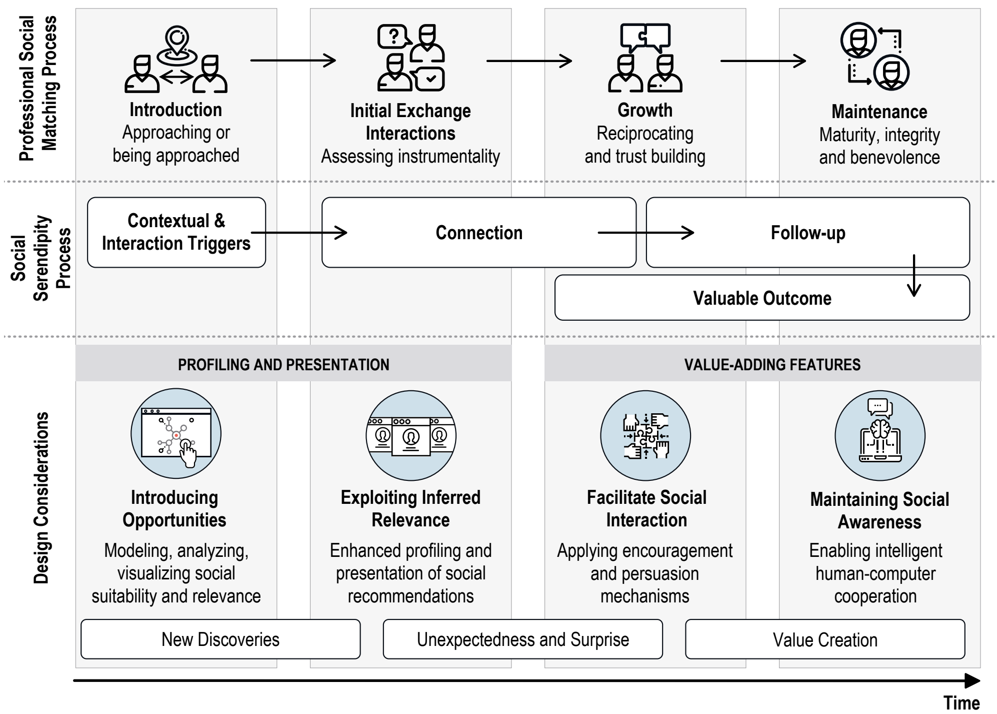


Figure 5.1 Overview of design consideration for serendipity-inducing and diversity-enhancing strategies in relation to social networking and serendipity processes.

Shifting social introductions to virtual environments might be affected by some fundamental limitations. For example, privacy constraints can prevent people from establishing new connections and reduce the desire to talk to strangers. The lack of exploratory, open-minded social settings in the virtual context could prevent people from spending enough time examining networking opportunities and shared interests. Therefore, future systems should be context-aware in defining user approachability while recommending new connections. As the presented findings suggest, one's approachability may vary depending on their environmental, personal, and social contexts. Therefore, it is essential to indicate contextual settings or opportune situations in which recommendations could be made. Although J. M. Mayer et al. (2015) have started work in this direction, technological means are needed to proactively identify opportune people to meet, opportune moments of encounter, and suitable and surprising ways to present suggestions to the user. Computational solutions could be a powerful tool to motivate users to act on social recommendations and increase their success rate when encountering new contacts.

In the following, I group the design considerations into two primary categories: (**i**) suggestions for enhancing the presentation of social recommendations and profiling; (**ii**) suggestions for value-adding features for encouraging and persuading social interactions. To exemplify the proposed consideration, I present visualizations of speculative design artifacts designed throughout the doctoral research.

5.2.1 Profiling and Presentation

Technology can facilitate the identification of relevant dimensions for PSM to inform the user data required for modeling and profiling. The interfaces for computational PSM are typically driven by the information acquired from people. Thus, the probability of experiencing unexpected social encounters depends on the content, such as other users' profiles and their presentation. The profile and additional information shown can serve as contextual triggers, initiating the introduction phase. How surprising and relevant the new contacts are can affect other users' acceptance of and willingness to interact with them.

Enriching the content. Self-created profile content will likely remain a central element in professional social matching. However, in virtual communication channels, people are limited in terms of non-verbal communication and counterbalance it by "sharing oneself" (Albrechtslund, 2008) through mediated content. It is necessary to investigate what types of self-representation knowledge workers consider valuable to start interacting with unfamiliar individuals and how to collect it unobtrusively. At the same time, profile data should go beyond professional details: Our findings hint that meaningful connections might also start from sharing non-work-related interests. To address this aspect, algorithmic user modeling and analysis of various explicit and implicit social data might allow us to efficiently derive latent insights regarding potential collaborators' social suitability and relevance. Data on individuals on social media sites may help to develop more in-depth and informative user profiles and identify relevant new and weak ties in the global social network (Olshannikova et al., 2020b). A suitable approach to enrich user profiles with highly semantic data is through interactions with virtual agent-assistants, as presented in the example of speculative service design in Figure 5.2 (A, B). User profiles should deliver content that guides the reader's attention to timely and relevant qualities, thereby initiating social interactions (see the example in Figure 5.2 C, D). Enriching the profiling can facilitate the recognition of inadvertent possibilities, implicit personal values, and contextual oddities (J. M. Mayer et al., 2016), characteristics that cannot be derived from user-generated material.

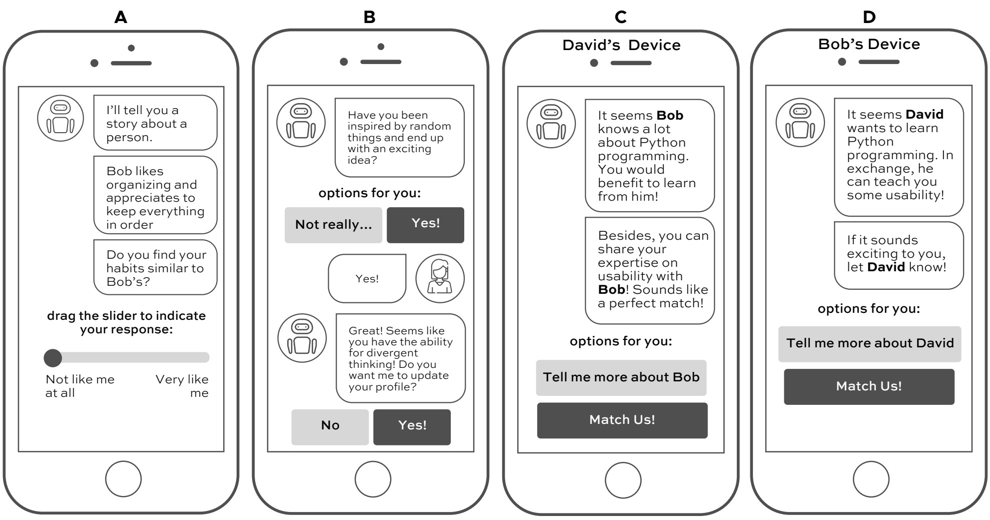


Figure 5.2 A & B: Examples of collecting rich semantic data on the user for enhanced profiling via a virtual agent; C & D: An example of nudging social interactions between users by introducing relevant facts on expertise and interests.

Facilitating inferred relevance with interactive visualization techniques. To promote serendipity, it is essential that users are in an opportunistic mindset. Visualizing opportunities in social graphs is a promising approach to enable fruitful social matching (Sen et al., 2011; Terveen & McDonald, 2005) and divergent search activities. Holistic, interactive, exploration-supporting visualizations can provide several perspectives to a broad range of future collaborators, exposing potential weak ties rather than people one already knows. Such a representation can serve as a trigger for unexpected discoveries, conveying cues about relevancy that can spark the users' interest and curiosity to evaluate suggested contacts. Social graphs could also provide visual cues on how diverse or similar recommendations are, depending on the need for partnering (see the example in Figure 5.3). In studies, some examples of social serendipity occurred between relatively like-minded people, while others were between people of different backgrounds and social circles. In a system, the optimal level of diversity is challenging to infer when a user is not actively seeking new connections. Therefore, there is a need for suitable methods for modeling a user's goals and aspirations, their entire social network, and their topical objectives and work areas. This would indicate the inferred relevance level of a prospective connection and communicate the expected contexts in which specific social recommendations were considered relevant.

5.2.2 Value-adding Features

The studies illustrated that encounters in both the physical and digital environments would benefit from facilitators transitioning from the initiation phase to follow-up activities and might require coordination to maintain relationships. Designing ways to motivate the user to follow up on the recommendations is thus an essential goal for future research. Particularly in PSM, the decision-making about whether to connect should be facilitated because opportunities can become numerous rapidly.

The uniqueness of computational support for PSM is defined by its proactivity (making opportune suggestions) and multidimensional algorithmic user modeling (beyond identifying similarity). In contrast to item recommender systems, the mechanism of social suitability and relevance is significantly different: A person who is a perfect match for one individual would not necessarily be relevant for another who shares similar qualities. Thus, existing people recommendation strategies often face challenges in converting the recommendations to user behavior due to the multifaceted nature of what constitutes a good social recommendation. There are also challenges related to translating factors related to personality and behavior into design choices because of the complexity of gathering, quantifying, and measuring such highly semantic data.

Designing ways to encourage users to discover and experience potential social serendipity calls for audacious design solutions. The persuasive and encouragement

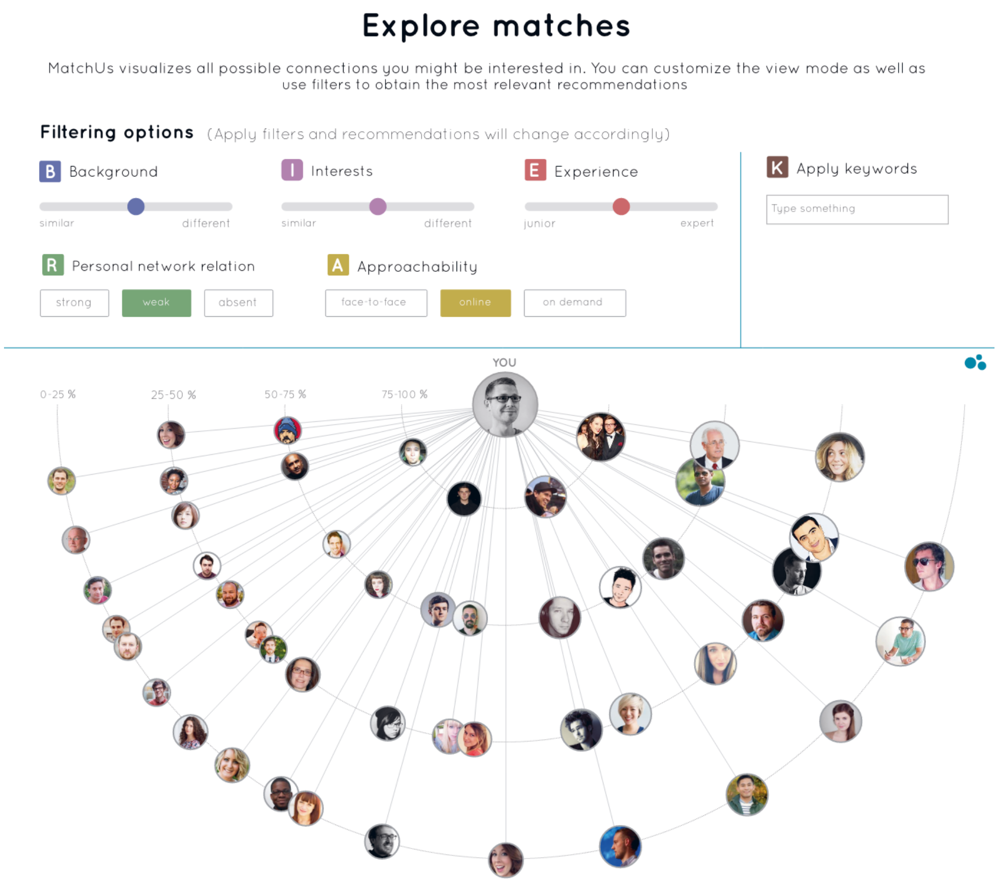


Figure 5.3 Example of utilizing interactive visualization for exploration of people recommendations. The user can adjust recommendations using various filters/dimensions such as interest and background diversity, seniority, structural position in the social network, and approachability as a geographical factor. Visualization thus provides clues about relevance from the perspective of the dimensions in use.

features could be achieved by employing a human-technology cooperation strategy. While technology excels in systematic, large-scale, big data-driven analytics, human beings still perform better in heuristic, intuition-based analyses, sense-making, and interpretation of qualitative aspects of social matching. Combining human and machine abilities is a vital HCI problem, which could augment intelligence (Corrigan, 2012; Zheng et al., 2017) by enabling a synergy of human and computational capabilities. Instead of replicating social serendipity qualities in digital environments, activating individuals' qualities responsible for the perception of serendipity by altering the use of technology is more relevant.

For example, technology can assist individuals by connecting the represented content and their needs, interests, or background (the connection phase of serendipity). It can also provide hints by inferring how a given recommendation is relevant to the target user, thus encouraging direct social interactions for value creation and providing the users with tickets-to-talk to initiate discussion. Additionally, to support the growth phase in PSM, timely notifications about the recommended person's recent activities, career updates, and topics of interest can facilitate knowledge of the connection's instrumental value and proactively encourage follow-up. Considering the human cognitive limitations, the technology can utilize storytelling approaches or adaptive and context-aware profiling to slowly disclose more information about connections (enabling the element of surprise). Such techniques could mimic repeated-encounter triggers revealed in the sample data, enhancing the decision-making process for value-creation opportunities.

5.3 Limitations

Naturally, multidisciplinary, exploratory, and experimental research has limitations that can affect the validity and reliability of the findings. Regarding generalizability, respondents mostly represent the same geographical area and cultural background due to the selected focus on academic circles. Long-term, large-scale studies may be needed to represent a broader demographic audience.

Proposed recommendation strategies focused only on different levels of topical and networking diversity, excluding important contextual PSM factors (e.g., complementarity of professional skills) that influence decision-making in selecting collaborators. While it can be seen as a limitation, prior research analysis revealed a recognized gap in the literature regarding the effectiveness of diversity-enhancing strategies in PSM. Before introducing more complex contextual factors to diversity-enhancing strategies, it was essential to establish a strong foundational understanding of if and how diversity, both in terms of topics and network connections, impacts PSM.

Methodologically, we had to control experiments to prove the effectiveness of the proposed diversity-enhancing strategies. The assessment of the reliability of the recommendation strategies remains quite preliminary. In future studies, it would be valuable to compare the proposed strategies' performance with traditional recommendation approaches to assess their effectiveness over existing standards. Additionally, driven by the experiment setup, the restrictions on user eligibility criteria and non-disclosure regarding how recommendations were produced could have influenced the evaluation. In future research, they can be disregarded and might yield new insights regarding the usefulness of the proposed strategies.

Furthermore, only three diversification calibrations were tested within the proposed strategies with a sample size limited by practicality and data availability. One of the crucial data limitations is that we did not filter out users with narrow online personas, such as junior researchers in Publication IV and users with fewer tweets or connections in Publication V. This factor could have limited the data analysis and decreased the accuracy of the individuals' representation, which in turn could affect the subjective perceptions of the recommendations. However, we intentionally wanted to introduce such recommendations to reveal how they could be appreciated in different PSM contexts and evaluate the role of seniority as a possible factor in collaboration processes.

Finally, addressing the effectiveness of proposed design considerations in significantly augmenting the role of technology in fostering serendipitous experiences presents a notable challenge, primarily due to the need for long-term impact assessments. Serendipity, which is a nuanced and often unpredictable phenomenon, requires an extended observation period to determine the lasting effects of technological interventions. Furthermore, the findings can have a potential bias within, where individuals may not recall instances where technology played a role in their serendipitous experiences as vividly as the examples they reported. This could potentially skew conclusions towards the perception that technology has a minimal role in facilitating serendipity. However, it is crucial to interpret this with caution. Acknowledging that technology might play a limited role should not translate into a categorical assertion that it should not be a factor in the serendipity equation. Instead, it prompts a need for a nuanced understanding of how technology can be designed and integrated effectively to augment rather than hinder serendipitous encounters, thereby underscoring the importance of future research in this domain.

Nevertheless, this dissertation lays the groundwork for new diversity-enhancing strategies in PSM, and it is essential to show that such alternative strategies are sensible from the users' viewpoint and technically feasible before comparing them with others. We call for follow-up research to compare the effectiveness of current and other alternative algorithmic approaches whose goal is diversity exposure in PSM.

5.4 Future Work

While this study has contributed valuable insights into understanding the role of diversity and serendipity in the design of diversity-enhancing strategies for PSM, several avenues for future research warrant further investigation. The following section highlights potential areas of exploration that can build upon the findings of this study and advance the research topic.

5.4.1 Replication, longitudinal, and comparative studies.

Conducting replication, longitudinal, and comparative studies can provide a more comprehensive understanding of how diversity-enhancing strategies impact PSM. These studies enable evaluating proposed strategies in relation to alternatives, assessing their suitability across different contexts and user groups, and determining their relative effectiveness in achieving research objectives.

Replication studies are vital for verifying and validating the findings of this research. First, future research can focus on replicating the experiments across different user groups to assess whether the strategies consistently lead to more diverse connections. Second, replicating the experiments on multiple social matching platforms or in various professional contexts can help determine whether the strategies hold true across different environments and user populations. Replication studies also enhance the generalizability of findings beyond a specific platform or setting.

Longitudinal studies can help to assess the long-term effects and outcomes of the variables examined in this study. Longitudinal designs would allow the examination of changes and patterns in the PSM process over time, providing a more comprehensive understanding of the dynamic nature of the diversity-exposure phenomena under investigation. For instance, longitudinal studies can help track the evolution of users' professional networks over time.

Conducting comparative studies of different groups or populations could shed light on the variations and similarities in the subjective perceptions of proposed diversity-enhancing strategies. Comparing groups with distinct personality characteristics or cultural backgrounds may uncover important nuances and contextual factors that influence the perception of relevance and willingness to follow up on recommendations. Such studies can also compare the sustained impact of diversity-enhancing strategies with that of conventional methods, shedding light on the long-term benefits and effects.

5.4.2 Mediating and moderating variables.

This study focused on a specific set of variables. However, future research could explore additional mediating and moderating variables that may influence the relationships this study identified. Investigating these factors could enhance the theoretical framework and contribute to a more nuanced understanding of the research topic. One potential mediating variable could be the degree of serendipity experienced during PSM interactions. Future research can investigate whether the proposed diversity-enhancing strategies influence the perception of serendipitous encounters. For instance, do these strategies increase the likelihood of users feeling that they've stumbled upon valuable connections they didn't anticipate? Does serendipity mediate the relationship between diversity-enhancing strategies and the perceived success of a social match? Another mediating variable could be the users' intentions and goals when engaging in PSM. You might explore whether users who approach social matching with a deliberate intention to diversify their networks are more likely to perceive the recommendations as effective. Does the intention to broaden one's professional horizons mediate the relationship between the strategies employed and the outcomes of social matching? Serendipity tolerance, or a user's openness to unexpected and fortunate encounters, can also be a moderating variable. Investigate whether individuals with a higher serendipity tolerance are more receptive to diversity-enhancing strategies. Do users with a greater willingness to embrace serendipity perceive the recommendations differently than those who prefer predictability in their networking efforts?

5.4.3 Ethical considerations.

It is incumbent upon designers and researchers to continue exploring and innovating in ways that harness the potential of diversity and serendipity while upholding ethical principles and ensuring that technology serves as an enabler of positive, inclusive, and equitable professional interactions. The emphasis on diversity exposure calls for careful attention to avoid perpetuating biases, as diversity should not merely be cosmetic but meaningful and inclusive. Future design efforts must prioritize fairness and equity, ensuring that the recommendations mechanisms do not inadvertently favor certain groups or reinforce existing inequalities. Additionally, striking the balance between personalized recommendations and opportunistic discoveries is crucial to ensure that technology enhances, rather than diminishes, the sense of agency and exploration in professional networking.

In summary, the future work section proposes several directions for future research that build upon the findings of this study. Longitudinal studies, replication studies, qualitative investigation, intervention studies, comparative studies, technological advancements, and further exploration of mediating and moderating variables all offer promising avenues for expanding knowledge and contributing to the advancement of the field. The findings underscore the responsibility to translate insights into design practices prioritizing diversity, transparency, and user agency. By doing so, we can navigate the complex landscape of PSM ethically, enhancing its potential to foster meaningful connections and collaborations in an increasingly interconnected world.

REFERENCES

Adamopoulos, P., & Tuzhilin, A. (2014). On unexpectedness in recommender systems: Or how to better expect the unexpected. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *5*(4), 1–32.

Agarwal, N. K. (2015). Towards a definition of serendipity in information behaviour. *Information research: an international electronic journal*, *20*(3), n3.

Agrawal, R., Gollapudi, S., Halverson, A., & Ieong, S. (2009). Diversifying search results. *Proceedings of the second ACM international conference on web search and data mining*, 5–14.

Akiyama, K., Kumamoto, T., & Nadamoto, A. (2017). Emotion-based method for latent followee recommendation in twitter. *Proc. of the 19th International Conference on Information Integration and Web-based Applications & Services*, 121–125.

Albrechtslund, A. (2008). Online social networking as participatory surveillance. *First Monday*, *13*(3).

Al-Taie, M. Z., Kadry, S., & Obasa, A. I. (2018). Understanding expert finding systems: Domains and techniques. *Social Network Analysis and Mining*, *8*(1), 57.

Amal, S., Tsai, C.-H., Brusilovsky, P., Kuflik, T., & Minkov, E. (2019). Relational social recommendation: Application to the academic domain. *Expert Systems with Applications*, *124*, 182–195.

Anderson, T. D. (2011). Beyond eureka moments: Supporting the invisible work of creativity and innovation. *Information Research: An international electronic journal*, *16*(1), n1.

André, P., Schraefel, M., Teevan, J., & Dumais, S. T. (2009). Discovery is never by chance: Designing for (un) serendipity. *Proceedings of the seventh ACM conference on Creativity and cognition*, 305–314.

Archer, M., & Zytko, D. (2019). Social matching systems for research collaboration: A profile page design for university faculty. *Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing*, 146–150. https://doi.org/10.1145/3311957.3359459

Arens-Volland, A., & Naudet, Y. (2016). Personalized recommender system for event attendees. *2016 11th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP)*, 65–70.

Bardzell, J. (2013). Critical and cultural approaches to hci. *The SAGE handbook of digital technology research*, 130.

Barlow, D. H., Ellard, K. K., Sauer-Zavala, S., Bullis, J. R., & Carl, J. R. (2014). The origins of neuroticism. *Perspectives on Psychological Science*, *9*(5), 481–496.

Bawden, D. (2011). Encountering on the road to serendip? browsing in new information environments. *Innovations in IR: Perspectives for theory and practice*, 1–22.

Beel, J., Gipp, B., Langer, S., & Breitinger, C. (2016). Research-paper recommender systems: A literature survey. *International Journal on Digital Libraries*, *17*(4, November 2016), 305–338.

Beel, J., Langer, S., Genzmehr, M., Gipp, B., Breitinger, C., & Nürnberger, A. (2013). Research paper recommender system evaluation: A quantitative literature survey. *Proceedings of the International Workshop on Reproducibility and Replication in Recommender Systems Evaluation*, 15–22.

Björneborn, L. (2017). Three key affordances for serendipity. *Journal of Documentation*.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, *3*(Jan), 993–1022.

Bobadilla, J., Ortega, F., Hernando, A., & Gutiérrez, A. (2013). Recommender systems survey. *Knowledge-based systems*, *46*, 109–132.

Bradley, K., & Smyth, B. (2001). Improving recommendation diversity. *Proceedings of the Twelfth Irish Conference on Artificial Intelligence and Cognitive Science, Maynooth, Ireland*, 85–94.

Brusilovsky, P., Oh, J. S., López, C., Parra, D., & Jeng, W. (2017). Linking information and people in a social system for academic conferences. *New Review of Hypermedia and Multimedia*, *23*(2), 81–111.

Buchem, I. (2011). Serendipitous learning: Recognizing and fostering the potential of microblogging. *Form@ re-Open Journal per la formazione in rete*, *11*(74), 7–16.

Buder, J., & Schwind, C. (2012). Learning with personalized recommender systems: A psychological view. *Computers in Human Behavior*, *28*(1), 207–216.

Cañamares, R., Castells, P., & Moffat, A. (2020). Offline evaluation options for recommender systems. *Information Retrieval Journal*, *23*(4), 387–410.

Carbonell, J., & Goldstein, J. (1998). The use of mmr, diversity-based reranking for reordering documents and producing summaries. *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*, 335–336.

Carraro, D., & Bridge, D. (2020). Debiased offline evaluation of recommender systems: A weighted-sampling approach. *Proceedings of the 35th Annual ACM Symposium on Applied Computing*, 1435–1442.

Carullo, G., Castiglione, A., De Santis, A., & Palmieri, F. (2015). A triadic closure and homophily-based recommendation system for online social networks. *World Wide Web*, *18*(6), 1579–1601.

Champiri, Z. D., Mujtaba, G., Salim, S. S., & Chong, C. Y. (2019). User experience and recommender systems. *2019 2nd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, 1–5.

Chin, A., Xu, B., Zhao, C., & Wang, X. (2014). From offline to online: Connecting people with a mobile social networking application at a conference. *Proceedings of the Second International Symposium of Chinese CHI*, 40–49.

Claes, R., & Ruiz-Quintanilla, S. A. (1998). Influences of early career experiences, occupational group, and national culture on proactive career behavior. *Journal of Vocational Behavior*, *52*(3), 357–378.

Clarke, C. L., Kolla, M., Cormack, G. V., Vechtomova, O., Ashkan, A., Büttcher, S., & MacKinnon, I. (2008). Novelty and diversity in information retrieval evaluation. *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, 659–666.

Collins, C. J., & Clark, K. D. (2003). Strategic human resource practices, top management team social networks, and firm performance: The role of human resource practices in creating organizational competitive advantage. *Academy of management Journal*, *46*(6), 740–751.

Corrigan, J. M. (2012). Augmented intelligence—the new ai—unleashing human capabilities in knowledge work. *2012 34th International Conference on Software Engineering (ICSE)*, 1285–1288.

Cremonesi, P., Garzotto, F., Negro, S., Papadopoulos, A. V., & Turrin, R. (2011). Looking for "good" recommendations: A comparative evaluation of recommender systems. *IFIP Conference on Human-Computer Interaction*, 152–168.

Cremonesi, P., Garzotto, F., & Turrin, R. (2013). User-centric vs. system-centric evaluation of recommender systems. *Ifip conference on human-computer interaction*, 334–351.

Cross, R., & Thomas, R. (2011). A smarter way to network. *Harvard Business Review*, *89*(7-8), 149–53.

Dautenhahn, K., & Ghauoi, C. (2014). The encyclopedia of human-computer interaction.

Davenport, T. H. (2005). *Thinking for a living: How to get better performances and results from knowledge workers*. Harvard Business Press.

De Vos, A., De Clippeleer, I., & Dewilde, T. (2009). Proactive career behaviours and career success during the early career. *Journal of Occupational and Organizational Psychology*, *82*(4), 761–777.

de Melo, R. M. C. (2018). *On serendipity in the digital medium: Towards a framework for valuable unpredictability in interaction design* (Doctoral dissertation). Faculdade de Belas Artes, Universidade do Porto. Porto, Portugal.

Dong, W., Ehrlich, K., Macy, M. M., & Muller, M. (2016). Embracing cultural diversity: Online social ties in distributed workgroups. *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, 274–287.

Eagle, N., & Pentland, A. (2005). Social serendipity: Mobilizing social software. *IEEE Pervasive Computing*, *4*(2), 28–34.

e Cunha, M. P., Clegg, S. R., & Mendonça, S. (2010). On serendipity and organizing. *European Management Journal*, *28*(5), 319–330.

Edizel, B., Bonchi, F., Hajian, S., Panisson, A., & Tassa, T. (2020). Fairecsys: Mitigating algorithmic bias in recommender systems. *International Journal of Data Science and Analytics*, *9*(2), 197–213.

Erickson, T., & Kellogg, W. A. (2000). Social translucence: An approach to designing systems that support social processes. *ACM transactions on computer-human interaction (TOCHI)*, *7*(1), 59–83.

Fleder, D. M., & Hosanagar, K. (2007). Recommender systems and their impact on sales diversity. *Proceedings of the 8th ACM conference on Electronic commerce*, 192–199.

for Standardization, I. O. (2010). *Ergonomics of human-system interaction: Part 210: Human-centred design for interactive systems*. ISO.

Fuchs, C. (2017). *Social media: A critical introduction*. Sage.

Garcia-Gavilanes, R., & Amatriain, X. (2010). Weighted content based methods for recommending connections in online social networks. *Proceedings of the ACM Conference on Recommender systems*, 68–71.

Gosling, S. D., Rentfrow, P. J., & Swann Jr, W. B. (2003). A very brief measure of the big-five personality domains. *Journal of Research in personality*, *37*(6), 504–528.

Granovetter, M. (2018). *Getting a job: A study of contacts and careers*. University of Chicago press.

Granovetter, M. (1983). The strength of weak ties: A network theory revisited. *Sociological theory*, 201–233.

Gurini, D. F., Gasparetti, F., Micarelli, A., & Sansonetti, G. (2013). A sentiment-based approach to twitter user recommendation. *Proceedings of the 5th ACM RecSys Workshop on Recommender Systems and the Social Web*, 1–4.

Guy, I. (2018). People recommendation on social media. In *Social information access* (pp. 570–623). Springer.

Guy, I. (2015). Social recommender systems. In *Recommender systems handbook* (pp. 511–543). Springer.

Guy, I., Jacovi, M., Perer, A., Ronen, I., & Uziel, E. (2010). Same places, same things, same people?: Mining user similarity on social media. In *Cscw'10. proceedings of the 2010 acm conference on computer supported cooperative work, savannah, georgia, usa, 06 – 10 february 2010* (pp. 41–50). ACM.

Guy, I., & Pizzato, L. (2016). People recommendation tutorial. In *Proc. of the 10th acm conference on recommender systems* (pp. 431–432). ACM.

Guy, I., Ronen, I., & Wilcox, E. (2009). Do you know? recommending people to invite into your social network. *Proceedings of the 14th international conference on Intelligent user interfaces*, 77–86.

Guy, I., Ur, S., Ronen, I., Perer, A., & Jacovi, M. (2011). Do you want to know? recommending strangers in the enterprise. *Proceedings of the ACM 2011 conference on Computer supported cooperative work*, 285–294.

Hannon, J., Bennett, M., & Smyth, B. (2010). Recommending twitter users to follow using content and collaborative filtering approaches. *Proceedings of the fourth ACM conference on Recommender systems*, 199–206.

Hartson, R. (2003). Cognitive, physical, sensory, and functional affordances in interaction design. *Behaviour & information technology*, *22*(5), 315–338.

Heck, T. (2013). Combining social information for academic networking. In *Cscw '13. proceedings of the acm conference on computer supported cooperative work, san antonio, tx, usa, 23 – 27 february 2013* (pp. 1387–1398). ACM.

Helberger, N., Karppinen, K., & D'Acunto, L. (2018). Exposure diversity as a design principle for recommender systems. *Information, Communication & Society*, *21*(2), 191–207.

Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS)*, *22*(1), 5–53.

Himmelsbach, J., Schwarz, S., Gerdenitsch, C., Wais-Zech-mann, B., Bobeth, J., & Tscheligi, M. (2019). Do we care about diversity in human computer interaction: A comprehensive content analysis on diversity dimensions in research. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 490.

Iaquinta, L., De Gemmis, M., Lops, P., Semeraro, G., Filannino, M., & Molino, P. (2008). Introducing serendipity in a content-based recommender system. *2008 Eighth International Conference on Hybrid Intelligent Systems*, 168–173.

Iaquinta, L., de Gemmis, M., Lops, P., Semeraro, G., & Molino, P. (2010). Can a recommender system induce serendipitous encounters. *E-commerce*, 1–17.

Jarrahi, M. H. (2017). Social media, social capital, and knowledge sharing in enterprise. *IT Professional*, *20*(4), 37–45.

Jeffrey, P., & McGrath, A. (2000). Sharing serendipity in the workplace. In *Proceedings of the third international conference on collaborative virtual environments* (pp. 173–179). ACM.

Johnson, S. (2011). *Where good ideas come from: The seven patterns of innovation*. Penguin UK.

Kaminskas, M., & Bridge, D. (2016). Diversity, serendipity, novelty, and coverage: A survey and empirical analysis of beyond-accuracy objectives in recommender systems. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, *7*(1), 1–42.

Kaminskas, M., & Bridge, D. (2014). Measuring surprise in recommender systems. *Proceedings of the workshop on recommender systems evaluation: dimensions and design (Workshop programme of the 8th ACM conference on recommender systems).*

Kay, F. (2010). *Successful networking: How to build new networks for career and company progression*. Kogan Page Publishers.

Knijnenburg, B. P., Willemsen, M. C., Gantner, Z., Soncu, H., & Newell, C. (2012). Explaining the user experience of recommender systems. *User modeling and user-adapted interaction*, *22*(4), 441–504.

Knijnenburg, B. P., Willemsen, M. C., & Kobsa, A. (2011). A pragmatic procedure to support the user-centric evaluation of recommender systems. *Proceedings of the fifth ACM conference on Recommender systems*, 321–324.

Kobsa, A. (2007). Privacy-enhanced web personalization. In *The adaptive web* (pp. 628–670). Springer.

Koene, A., Perez, E., Carter, C. J., Statache, R., Adolphs, S., O'Malley, C., Rodden, T., & McAuley, D. (2015). Ethics of personalized information filtering. *International Conference on Internet Science*, 123–132.

Kong, X., Jiang, H., Yang, Z., Xu, Z., Xia, F., & Tolba, A. (2016). Exploiting publication contents and collaboration networks for collaborator recommendation. *PloS one*, *11*(2), e0148492.

Konstan, J. A., & Riedl, J. (2012). Recommender systems: From algorithms to user experience. *User modeling and user-adapted interaction*, *22*(1), 101–123.

Koprinska, I., & Yacef, K. (2015). People-to-people reciprocal recommenders. In *Recommender systems handbook* (pp. 545–567). Springer.

Kossinets, G., & Watts, D. J. (2009). Origins of homophily in an evolving social network. *American journal of sociology*, *115*(2), 405–450.

Kotkov, D., Wang, S., & Veijalainen, J. (2016). A survey of serendipity in recommender systems. *Knowledge-Based Systems*, *111*, 180–192.

Kunaver, M., & Požrl, T. (2017). Diversity in recommender systems–a survey. *Knowledge-Based Systems*, *123*, 154–162.

Lang, F. R., John, D., Lüdtke, O., Schupp, J., & Wagner, G. G. (2011). Short assessment of the big five: Robust across survey methods except telephone interviewing. *Behavior research methods*, *43*(2), 548–567.

Leino, J. (2014). User factors in recommender systems: Case studies in e-commerce, news recommending, and e-learning.

Li, B., & Han, L. (2013). Distance weighted cosine similarity measure for text classification. *International Conference on Intelligent Data Engineering and Automated Learning*, 611–618.

Lin, C.-Y., Cao, N., Liu, S. X., Papadimitriou, S., Sun, J., & Yan, X. (2009). Smallblue: Social network analysis for expertise search and collective intelligence. *2009 IEEE 25th International Conference on Data Engineering*, 1483–1486.

Lopes, G. R., Moro, M. M., Wives, L. K., & De Oliveira, J. P. M. (2010). Collaboration recommendation on academic social networks. *International conference on conceptual modeling*, 190–199.

Lutz, C., Ranzini, G., & Meckel, M. (2013). Trusted surprises?: Antecedents of serendipitous encounters online.

Makri, S., & Blandford, A. (2011). What is serendipity?-a workshop report. *Information Research*, *16*(3), 491.

Makri, S., Blandford, A., Woods, M., Sharples, S., & Maxwell, D. (2014). "making my own luck": Serendipity strategies and how to support them in digital information environments. *Journal of the Association for Information Science and Technology*, *65*(11), 2179–2194.

Mayer, J. (2014). Is there a place for serendipitous introductions? In *Proceedings of the companion publication of the 17th acm conference on computer supported cooperative work & social computing* (pp. 73–76). ACM.

Mayer, J., & Jones, Q. (2016). Encount'r: Exploring passive context-awareness for opportunistic social matching. *Proceedings of the 19th ACM Conference on*

Computer Supported Cooperative Work and Social Computing Companion, 349–352.

Mayer, J. M., Hiltz, S. R., Barkhuus, L., Väänänen, K., & Jones, Q. (2016). Supporting opportunities for context-aware social matching: An experience sampling study. In *Chi'16. proceedings of the acm conference on human factors in computing systems, san jose, california, usa, 07 – 12 may 2016* (pp. 2430–2441). ACM.

Mayer, J. M., Jones, Q., & Hiltz, S. R. (2015). Identifying opportunities for valuable encounters: Toward context-aware social matching systems. *ACM Transactions on Information Systems (TOIS)*, *34*(1), 1.

McBirnie, A., Ford, S., McCay-Peet, L., & Makri, S. R. (2016). Serendipity in future digital information environments. In T. M. & S. R. Makri (Eds.), *Accidental information discovery. cultivating serendipity in the digital age* (pp. 81–114). Elsevier.

McBirnie, A. (2008). Seeking serendipity: The paradox of control. *Aslib Proceedings: New Information Perspectives*, *60*(6), 600–618.

McCay-Peet, L., & Quan-Haase, A. (2016). The influence of features and demographics on the perception of twitter as a serendipitous environment. *Proceedings of the 27th ACM Conference on Hypertext and Social Media*, 333–335.

McCay-Peet, L., & Toms, E. G. (2015). Investigating serendipity: How it unfolds and what may influence it. *Journal of the Association for Information Science and Technology*, *66*(7), 1463–1476.

McCay-Peet, L., & Wells, P. G. (2017). Serendipity in the sciences–exploring the boundaries. *Proceedings of the Nova Scotian Institute of Science (NSIS)*, *49*(1), 97.

McGrenere, J., & Ho, W. (2000). Affordances: Clarifying and evolving a concept. In *Gi 2000. proceedings of graphics interface, montréal, québec, canada, 15-17 may 2000* (pp. 179–186). Canadian Human-Computer Communications Society.

McNee, S. M., Riedl, J., & Konstan, J. A. (2006). Being accurate is not enough: How accuracy metrics have hurt recommender systems. *CHI'06 extended abstracts on Human factors in computing systems*, 1097–1101.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual review of sociology*, 415–444.

Merton, R. K., & Barber, E. (2011). *The travels and adventures of serendipity: A study in sociological semantics and the sociology of science*. Princeton University Press.

Michael, J., & Yukl, G. (1993). Managerial level and subunit function as determinants of networking behavior in organizations. *Group & organization management*, *18*(3), 328–351.

Mitchell, R., & Nicholas, S. (2006). Knowledge creation in groups: The value of cognitive diversity, transactive memory and open-mindedness norms. *The Electronic Journal of Knowledge Management (EJKM)*, *4*(1, December 2006), 67–74.

Muller, M., Ehrlich, K., Matthews, T., Perer, A., Ronen, I., & Guy, I. (2012). Diversity among enterprise online communities: Collaborating, teaming, and innovating through social media. In *Chi'12. proceedings of the sigchi conference on human factors in computing systems, austin, texas, usa 5 – 10 may 2012* (pp. 2815–2824). ACM.

Nguyen, C. T. (2020). Echo chambers and epistemic bubbles. *Episteme*, *17*(2), 141–161.

Nguyen, T. T., Hui, P.-M., Harper, F. M., Terveen, L., & Konstan, J. A. (2014). Exploring the filter bubble: The effect of using recommender systems on content diversity. *Proceedings of the 23rd international conference on World wide web*, 677–686.

Nutefall, J. E., & Ryder, P. M. (2010). The serendipitous research process. *The Journal of Academic Librarianship*, *36*(3), 228–234.

Oku, K., & Hattori, F. (2011). Fusion-based recommender system for improving serendipity. *DiveRS@ RecSys*, *816*, 19–26.

Olshannikova, E., Olsson, T., Huhtamäki, J., Paasovaara, S., & Kärkkäinen, H. (2020a). From chance to serendipity: Knowledge workers' experiences of serendipitous social encounters. *Advances in Human-Computer Interaction*, *2020*(1827107), 1–18. https://doi.org/https://doi.org/10.1155/2020/1827107

Olsson, T., Huhtamäki, J., & Kärkkäinen, H. (2020). Directions for Professional Social Matching Systems. *Communications of the ACM*, *63*(2), 60–69. https://doi.org/10.1145/3363825

Ookalkar, R., Reddy, K. V., & Gilbert, E. (2019). Pop: Bursting news filter bubbles on twitter through diverse exposure. *Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing*, 18–22.

Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, *9*(1), 62–66.

Paasovaara, S., Olshannikova, E., Jarusriboonchai, P., Malapaschas, A., & Olsson, T. (2016). Next2you: A proximity-based social application aiming to encourage interaction between nearby people. *Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia*, 81–90.

Parise, S., Whelan, E., & Todd, S. (2015). How twitter users can generate better ideas. *MIT Sloan Management Review*, *56*(4), 21.

Pariser, E. (2011). *The filter bubble: What the internet is hiding from you*. Penguin UK.

Pease, A., Colton, S., Ramezani, R., Charnley, J. W., & Reed, K. (2013). A discussion on serendipity in creative systems. *ICCC*, 64–71.

Pizzato, L., Rej, T., Akehurst, J., Koprinska, I., Yacef, K., & Kay, J. (2013). Recommending people to people: The nature of reciprocal recommenders with a case study in online dating. *User Modeling and User-Adapted Interaction*, *23*(5), 447–488.

Porter, C. M., & Woo, S. E. (2015). Untangling the networking phenomenon: A dynamic psychological perspective on how and why people network. *Journal of Management*, *41*(5), 1477–1500.

Pu, P., Chen, L., & Hu, R. (2012). Evaluating recommender systems from the user's perspective: Survey of the state of the art. *User Modeling and User-Adapted Interaction*, *22*(4), 317–355.

Pu, P., Chen, L., & Hu, R. (2011). A user-centric evaluation framework for recommender systems. *Proceedings of the fifth ACM conference on Recommender systems*, 157–164.

Resnick, P., Garrett, R. K., Kriplean, T., Munson, S. A., & Stroud, N. J. (2013). Bursting your (filter) bubble: Strategies for promoting diverse exposure. *Proceedings of the 2013 conference on Computer supported cooperative work companion*, 95–100.

Robert, L. P. (2016). Far but near or near but far?: The effects of perceived distance on the relationship between geographic dispersion and perceived diversity. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2461–2473.

Robison, M. K., Gath, K. I., & Unsworth, N. (2017). The neurotic wandering mind: An individual differences investigation of neuroticism, mind-wandering, and executive control. *The Quarterly Journal of Experimental Psychology*, *70*(4), 649–663.

Sanz-Cruzado, J., & Castells, P. (2018). Enhancing structural diversity in social networks by recommending weak ties. *Proceedings of the 12th ACM Conference on Recommender Systems*, 233–241.

Sen, S., Charlton, H., Kerwin, R., Lim, J., Maus, B., Miller, N., Naminski, M. R., Schneeman, A., Tran, A., Nunes, E., et al. (2011). Macademia: Semantic visualization of research interests. *Proceedings of the 16th international conference on Intelligent user interfaces*, 457–458.

Shah, K., Salunke, A., Dongare, S., & Antala, K. (2017). Recommender systems: An overview of different approaches to recommendations. *2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, 1–4.

Sillence, E., & Briggs, P. (2007). Please advise: Using the internet for health and financial advice. *Computers in Human Behavior*, *23*(1), 727–748.

Sinha, R., & Swearingen, K. (2002). The role of transparency in recommender systems. *CHI'02 extended abstracts on Human factors in computing systems*, 830–831.

Sîrbu, A., Pedreschi, D., Giannotti, F., & Kertész, J. (2019). Algorithmic bias amplifies opinion fragmentation and polarization: A bounded confidence model. *PloS one*, *14*(3).

Smyth, B., & McClave, P. (2001). Similarity vs. diversity. *International conference on case-based reasoning*, 347–361.

Sparrowe, R. T., Liden, R. C., Wayne, S. J., & Kraimer, M. L. (2001). Social networks and the performance of individuals and groups. *Academy of management journal*, *44*(2), 316–325.

Sun, X., Sharples, S., & Makri, S. (2011). A user-centred mobile diary study approach to understanding serendipity in information research. *Information Research*, *16*(3), 16–3.

Terren, L., & Borge-Bravo, R. (2021). Echo chambers on social media: A systematic review of the literature. *Review of Communication Research*, *9*, 99–118.

Terveen, L., & McDonald, D. W. (2005). Social matching: A framework and research agenda. *ACM transactions on computer-human interaction (TOCHI)*, *12*(3), 401–434.

Tsai, C.-H., & Brusilovsky, P. (2019). Exploring social recommendations with visual diversity-promoting interfaces. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, *10*(1), 1–34.

Tsai, C.-H., & Brusilovsky, P. (2016). A personalized people recommender system using global search approach. *IConference 2016 Proceedings*.

Van Hoye, G., & Saks, A. M. (2008). Job search as goal-directed behavior: Objectives and methods. *Journal of Vocational Behavior*, *73*(3), 358–367.

Van Le, T., Truong, T. N., & Pham, T. V. (2014). A content-based approach for user profile modeling and matching on social networks. *International Workshop on Multi-disciplinary Trends in Artificial Intelligence*, 232–243.

Vargas, S., Baltrunas, L., Karatzoglou, A., & Castells, P. (2014). Coverage, redundancy and size-awareness in genre diversity for recommender systems. *Proceedings of the 8th ACM Conference on Recommender systems*, 209–216.

Vargas, S., & Castells, P. (2011). Rank and relevance in novelty and diversity metrics for recommender systems. *Proceedings of the fifth ACM conference on Recommender systems*, 109–116.

Wang, H.-C., Fussell, S. R., & Cosley, D. (2011). From diversity to creativity: Stimulating group brainstorming with cultural differences and conversationally-retrieved pictures. *Proceedings of the ACM 2011 conference on Computer supported cooperative work*, 265–274.

Wolff, H.-G., Moser, K., & Grau, A. (2008). Networking: Theoretical foundations and construct validity. *Readings in applied organizational behavior from the Lüneburg Symposium*, 101–118.

Xiao, B., & Benbasat, I. (2007). E-commerce product recommendation agents: Use, characteristics, and impact. *MIS quarterly*, 137–209.

Yogev, A., Guy, I., Ronen, I., Zwerdling, N., & Barnea, M. (2015). Social media-based expertise evidence. *ECSCW 2015: Proceedings of the 14th European Conference on Computer Supported Cooperative Work, 19-23 September 2015, Oslo, Norway*, 63–82.

Yuan, Q., Cong, G., Zhao, K., Ma, Z., & Sun, A. (2015). Who, where, when, and what: A nonparametric bayesian approach to context-aware recommendation and search for twitter users. *ACM Transactions on Information Systems (TOIS)*, *33*(1), 2.

Yuan, Y. C., & Gay, G. (2006). Homophily of network ties and bonding and bridging social capital in computer-mediated distributed teams. *Journal of computer-mediated communication*, *11*(4), 1062–1084.

Yukl, G. (2012). Effective leadership behavior: What we know and what questions need more attention. *Academy of Management Perspectives*, *26*(4), 66–85.

Zhang, A. X., Bhardwaj, A., & Karger, D. (2016). Confer: A conference recommendation and meetup tool. *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion*, 118–121.

Zhang, L., Fang, H., Ng, W. K., & Zhang, J. (2011). Intrank: Interaction ranking-based trustworthy friend recommendation. *2011IEEE 10th International Conference on Trust, Security and Privacy in Computing and Communications*, 266–273.

Zhang, Y. C., Séaghdha, D. Ó., Quercia, D., & Jambor, T. (2012). Auralist: Introducing serendipity into music recommendation. *Proceedings of the fifth ACM international conference on Web search and data mining*, 13–22.

Zheng, N.-n., Liu, Z.-y., Ren, P.-j., Ma, Y.-q., Chen, S.-t., Yu, S.-y., Xue, J.-r., Chen, B.-d., & Wang, F.-y. (2017). Hybrid-augmented intelligence: Collaboration and cognition. *Frontiers of Information Technology & Electronic Engineering*, *18*(2), 153–179.

Ziegler, C.-N., McNee, S. M., Konstan, J. A., & Lausen, G. (2005). Improving recommendation lists through topic diversification. *Proceedings of the 14th international conference on World Wide Web*, 22–32.

Zytko, D., & DeVreugd, L. (2019). Designing a social matching system to connect academic researchers with local community collaborators. *Proc. ACM Hum.-Comput. Interact.*, *3*(GROUP). https://doi.org/10.1145/3361117

PUBLICATIONS

PUBLICATION
I

From chance to serendipity: Knowledge workers' experiences of serendipitous social encounters

Olshannikova, E., Olsson, T., Huhtamäki, J., Paasovaara, S., and Kärkkäinen, H.,

Advances in Human-Computer Interaction, *2020*(1827107), 1–18
https://doi.org/10.1155/2020/1827107

Hindawi
Advances in Human-Computer Interaction
Volume 2020, Article ID 1827107, 18 pages
https://doi.org/10.1155/2020/1827107

Research Article

From Chance to Serendipity: Knowledge Workers' Experiences of Serendipitous Social Encounters

Ekaterina Olshannikova,[1] **Thomas Olsson**,[1] **Jukka Huhtamäki**,[2] **Susanna Paasovaara**,[1] **and Hannu Kärkkäinen**[2]

[1]*Faculty of Information Technology and Communication Studies, Tampere University, Kalevantie 4, 33100 Tampere, Finland*
[2]*Faculty of Management and Business, Tampere University, Korkeakouluntie 8, 33720 Tampere, Finland*

Correspondence should be addressed to Ekaterina Olshannikova; ekaterina.olshannikova@tuni.fi

Received 7 October 2019; Revised 2 December 2019; Accepted 27 December 2019; Published 14 February 2020

Academic Editor: Francesco Bellotti

Serendipity refers to uncontrolled circumstances that lead to unexpected yet fortunate discoveries. The phenomenon has been studied extensively in relation to information retrieval. However, serendipity in the context of social encounters has been the subject of few empirical studies. In professional life, social serendipity might result in benefits such as fruitful collaboration, successful recruitment, discovery of novel information, and acquisition of crucial new perspectives from peers. Despite the potential significance of serendipity, particularly for knowledge work, there is a lack of empirical understanding of related subjective experiences and the role of technology within the process of encountering unsought findings. This qualitative study investigates knowledge workers' detailed narratives of serendipitous social encounters and the related factors through an analysis of 37 responses to an international online survey. We provide a detailed account of the experiential characteristics and contextual qualities of the reported instances of social serendipity. Finally, we discuss the seemingly minor role of technology in social serendipity and research avenues to computationally enhance social serendipity.

1. Introduction

Originating in the 18th century [1], the concept of serendipity has been researched as a phenomenon of uncontrolled circumstances that lead to unexpected yet fortunate discoveries [2]. Serendipity unfolds as the personal ability [3] to benefit from happy accidents [4]. The general notion of serendipity has been studied extensively in relation to creativity and innovation activities [5, 6], information retrieval [7, 8], and knowledge building and learning [9, 10]. Consequently, serendipity has often been conceptualized as chance or luck [11–13], particularly in the context of innovation processes and scientific discoveries.

Earlier research in human-computer interaction (HCI) and computer-supported cooperative work (CSCW) was designed to capture serendipity within established work environments, with the aim to facilitate intraorganizational knowledge creation and dissemination. As a result, two research approaches have emerged [14]: the exploration of natural serendipity (entirely unpredictable, nondeterministic, and nonfacilitated) and design for artificial serendipity (facilitated or triggered with the help of artificial agents such as information communication technology (ICT) applications). Natural serendipity has been approached in exploratory studies on daily, spontaneous encounters [15] and social awareness [16] in co-located work environments. Constructive research on artificial serendipity has focused on designing systems that enable chance encounters or so-called impromptu encounters [17], with the aim to enhance social awareness and interactions among collocated or distributed workers. In the context of information retrieval, a typical example of artificial serendipity is enabling surprising, novel discoveries in content-based recommender systems to improve the diversity of recommendations [18–20].

While much CSCW and HCI research has focused on understanding and supporting the elements of chance and surprise, we agree with theories stressing that chance encounters do not always lead to serendipitous events. The element of benefits is equally central, as emphasized by research on information search and ICT [21–23]. The subject of this article, therefore, is an exploration and analysis of how chance encounters turn into professionally valuable experiences of social serendipity (the example story in Figure 1), followed by a discussion of how ICT can better support this process. We focus on encounters between knowledge workers or people in professions that require high levels of creativity, extensive use of intellectual skills, and theoretical rather than contextual knowledge [24]. According to Davenport [25], knowledge workers possess *"high degrees of expertise, education, or experience and the primary purpose of their jobs involves the creation, distribution, or application of knowledge."*

Highly networked and collaborative, modern knowledge work can benefit from social serendipity because workers' tasks often require extensive social networks, various types of information, and access to people with complementary expertise [26]. It is noteworthy that, in today's knowledge work practices, value is often created in an ecosystemic way [27] and through social networks [28]. Organizational fluidity [29] relaxes conservative boundaries and structures, allowing networking and collaboration to take place more freely within and among organizations, cultures, and disciplines. Such changes in collaboration and networking practices have introduced new interest in fostering serendipitous encounters regardless of actors' affiliations and across organizational boundaries. We anticipate that such interorganizational social serendipity can manifest in, for example, forming new, useful connections at networking events and recruitment fairs, identifying relevant peers at conferences, and establishing fruitful business relationships in cocreation spaces and start-up incubators [30].

We argue that the experience of social serendipity that takes place naturally outside the workplace is an indicator of successful, desirable knowledge work. For individual, social serendipity can be a highly beneficial experience, providing both emotional pleasure and instrumental gain. Designing information systems for facilitating serendipity could benefit from deeper empirical understanding of what examples people have of such experiences and how social serendipity emerges as an experience. This knowledge, in turn, could shed light on the expected role of ICT in social serendipity processes and how technology can contribute to formation and flow of serendipitous experiences in professional life.

This research is driven by the following questions: *(RQ1) What characterizes serendipity in the context of social encounters among knowledge workers? (RQ2) What behavioral and experiential processes are associated with social serendipity?* Through an analysis of 37 knowledge workers' self-reported narrations of their experiences of serendipitous social encounters, we provide a detailed account of the various experiential and contextual qualities that impact social serendipity. Additionally, we extend the theoretical understanding of the traditional serendipity process with characteristics particularly applicable to social serendipity.

2. Related Work

This section first dives deeply into the concept of serendipity and describes relevant theories and studies to identify the research gaps we address in this article. Second, we review the existing research that envisions opportunities to artificially support serendipity via various technological solutions.

2.1. Theoretical Foundations of Serendipity. The conceptualization of serendipity began in 1754 through the story about "The Three Princes of Serendip" who were *"always making discoveries, by accidents and sagacity, of things they were not in quest of"* [31]. The term gained popularity in the late 20th century in a detailed discussion on serendipity as *"the art of making unsought findings"* [32]. Researchers theorizing serendipity were primarily interested in the role of chance or luck in scientific discoveries. An alternative definition appeared in the work of Liang [33], who studied the concept as an experiential quality within interactive technologies and defined it as a *"pure experience in our everyday lives interwoven with a tangled ecology of interactive systems."*

Many researchers have investigated serendipity in the domain of information retrieval and produced various frameworks and theories of its elements under different names despite major similarities in their meanings. According to Makri and Blandford [21], serendipity can be initiated by a conducive physical environment in the absence of time pressures. Other researchers labelled these features as triggers [22, 23]. Makri and Blandford [21] noticed that to benefit from serendipity, one should also have an implicit awareness of the need or opportunity for unsought discoveries and an open mind to sense-making of serendipitous cues. These observations are reflecting the notion of the prepared mind proposed by Pease et al. [22] and the element of connection in the model developed by McCay-Peet and Toms [23]. The final element common across all models of serendipity is the so-called serendipitous or valuable outcome, which might refer to the creation of a product, artifact, or knowledge.

In addition to these process-based models, Sun et al. [34] and Zhou et al. [35] developed a serendipity model based on three types of contexts that impact the serendipitous experience. First, the external context refers to the role of personal status and temporal and environmental factors, for example, encountering unsought findings during activities such as leisure and work in a specific place and at a certain time. Second, the social context relates to unexpected encounters achieved through socializing with both familiar and unfamiliar others. The external and social contexts thus facilitate perceptions of unexpectedness through various surrounding stimuli, reflecting the earlier conceptualizations of conducive physical environments and triggers. Finally, the internal context comprises the role of individuals' background knowledge, experiences, mindsets, recognized needs, emotional states, and levels of perceptiveness. This context is

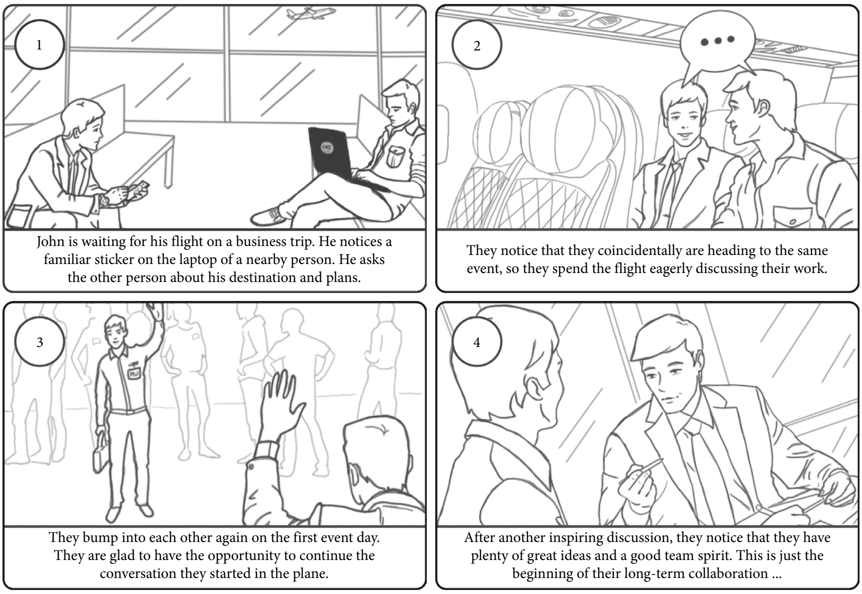


Figure 1: Fictional example of an unexpected encounter that turns into social serendipity.

tied to the sense-making of unexpected encounters and value creation. The focus on individuals' characteristics in this view on serendipity can be clearly associated with the notion of the prepared mind, while the sense-making element of the proposed model correlates with the notions of connection and valuable outcomes mentioned earlier.

In this article, we adopt McCay-Peet and Toms's [23] model as the central theoretical framework and analytic tool for the collected experiences of serendipity. Reviewing existing models that theorize serendipity as a process and propose a temporal framework also allows us to analytically dissect the experiences collected through the survey. McCay-Peet and Toms [23] distinguished the following phases of the serendipitous experience. The trigger initiates the serendipity experience and thought process about a potentially valuable outcome. The experience can arise from visual nontextual cues such as observations of activities and environments, verbal cues in conversations among individuals, and textual cues from reading content in books, journals, and websites. In the connection phase, a person relates the trigger with their background and knowledge, which leads to identification of a valuable outcome, defined as the potential to solve an existing problem or open new opportunities. The follow-up phase consists of the actions a person takes to achieve a valuable outcome such as capturing the trigger for later use, immediately acting on the opportunity, or preparing to accomplish a valuable outcome. McCay-Peet and Toms [23] also identified three key categories of valuable outcomes that have personal, organizational, and global effects. An unexpected thread is a sign that, when perceived by a person, indicates the presence of luck, chance, accident, or surprise in a serendipitous situation. We apply this model in the analysis of the reported experiences and explore its validity in the context of social serendipity in knowledge work.

In summary, according to earlier conceptualizations of serendipity, to take advantage of serendipity, one should have an open mood and free time to turn opportunities into action. A person should be able to perceive and analyze serendipitous cues from the environment to gain benefits. Following this broad theorization, we want to emphasize that experiences other than luck and chance must unfold to fulfil the requirements of the serendipity definitions. It is noteworthy that serendipity takes place in the realm of experiencing, which does not permit objective examination or measurement. Both essential elements of serendipity—unexpectedness and benefits—are primarily subjectively defined. The question of what is sufficiently unexpected or beneficial to count as serendipity varies among individuals. It, therefore, remains unclear how the natural contingency in everyday social encounters turns into something recognized as useful. Some long-term effects such as the increased productivity of individuals and organizations and the creation of new ideas can be recognized objectively. However, it is almost oxymoronic to empirically study the interplay between the experience and long-term effects of serendipity. It is difficult for individuals to distinguish between the experienced realm and physical reality [36], so rendering the scientific study of such concepts remains challenging. Perhaps due to this experiential nature, most theorizations of serendipity are limited to describing the internal and external factors that contribute to the probability of a particular event such as finding a new interesting connection. To our knowledge, particularly in our interest area of knowledge work, there are neither proper conceptualizations nor prior empirical studies on individuals' subjective experiences of the overall social serendipity process from the trigger to the valuable outcome.

2.2. Designing for Serendipity

2.2.1. Key Affordances for Artificial Serendipity. Serendipitous experiences triggered or initiated by physical and digital objects, agents, or environments have been termed artificial serendipity [14]. The concept also appears in the literature where it is called controlled [37] and online serendipity [38, 39]. In contrast to its natural counterpart, which is entirely unpredictable and nondeterministic, artificial serendipity happens only when necessary conditions occur. De Melo [14] classified four categories of triggers for serendipity (Table 1): things, places, events, and agents.

Such physical and digital triggers might possess key affordances for serendipity, as conceptualized by Björneborn [40]. The first affordance—so-called diversifiability—describes the capacity of environments to enable access to diverse, dissimilar, and incomplete contents. This affordance, coupled with individuals' ability to be open-minded, can spark curiosity and interest. The second affordance—traversability—refers to the environment capacity to enable individuals' mobility, leading to divergent exploration and convergent search activities. The most rudimentary examples of diversifiability and traversability affordances in the digital realm are the Google search engine and Twitter feeds, which increase the chances for serendipity to occur by opening access to enormous amounts of contents and different paths to encounter them. As an example of a physical environment with these two affordances, Björneborn [40] mentioned libraries. Finally, sensorability relates to the degree of stimuli richness in the environment and the capacity to stimulate the serendipitous experience through various senses. Coupled with individuals' sensitivity and attention, sensorability is responsible for producing a sense of surprise and experience. To summarize, diversifiability and traversability manifest in both physical and digital environments and have stronger effects in the latter due to the ease of accessing and manipulating information. Sensorability primarily manifests within things and trigger agents. Digital triggers typically are limited to visual and audio stimuli, so the sensorability affordance is prevalent within the physical realm involving all the senses.

Digital environments' potential to have affordances for serendipity was addressed by McCay-Peet and Quan-Haase [41], who argued that dynamic, diverse virtual environments can become engines of serendipity. The nature of dynamism and diversity as key qualities of digital environments can be correlated to the diversifiability and traversability affordances. McCay-Peet and Quan-Haase [41] studied users' perceptions of Twitter as a serendipitous environment to outline a design space for future serendipity-embedded digital services. The main finding from their research was that greater activity on such platforms can strengthen perceptions of serendipity and increase the probability of opportunistic information discovery (i.e., divergent content exploration). Additionally, the authors concluded that perceptions of serendipity might vary due to age differences and might be influenced by the various motivational factors behind the use of social platforms.

In contrast, Lutz et al. [42] challenged the possibility of serendipitous online experiences, arguing that algorithms might decrease the element of surprise in discovery. The authors demonstrated that Internet is perceived as a serendipitous engine due to the abundance of transparent information available for discovery. In addition, the authors confirmed that serendipity depends on trust and privacy, which can either prevent or establish the context for offline and online serendipity. They also found that people with mindsets prepared for opportune discoveries have fewer trust and privacy concerns. Thus, prior research has demonstrated the interplay of external affordances related to various physical and digital triggers and internal affordances related to individuals' characteristics and abilities.

2.2.2. Supporting Serendipitous Social Encounters. Although earlier CSCW and HCI research did not refer directly to the term "social serendipity," several concepts are somewhat related to the phenomenon. For example, Kiesler and Cummings [15] found that frequent, informal, spontaneous interactions in collocated work environments enable cohesive relationships among knowledge workers and increase their social awareness. Similarly, Vyas et al. [16] suggested that spontaneous encounters play significant roles in working life because employees in organizations rarely have opportunities for direct interactions that make them aware of others' activities. In particular, regarding the role of physical places that trigger such encounters, Brown et al. [43] concluded that the workspace layout may facilitate unplanned, casual interactions among employees and can serve as a contextual cue for enhanced communication and productivity in organizations.

Such exploratory studies have motivated various system designs to increase the probability of chance encounters and overall social awareness within organizations. For instance, Erickson and Kellogg [44] introduced the concept of socially translucent systems aimed at increasing the visibility of employees' activities in large groups and organizations. Jeffrey and McGrath [17] designed the "Forum," a collaborative working environment as a space for informal online interactions that help employees make new connections and share knowledge.

Studies on social matching have also been aimed at understanding the role of both personal (internal) and contextual (external) factors in supporting serendipitous encounters between strangers [45, 46]. Eagle and Pentland [47] proposed research on sensing contextual surroundings and increasing serendipitous interactions via mobile devices. They built an artifact called serendipity, a socially curious mobile device that encourages face-to-face interactions within a range of proximity. Interestingly, this solution is based on the idea of maximizing similarity in the matching process, which might lead to anticipated rather than serendipitous encounters.

Table 1: Examples of physical and digital triggers that can initiate artificial serendipity.

	Trigger type	Physical	Digital
What	*Things*	Inanimate material objects (e.g., pens, desks, cups, desks, and plants)	Digital artifacts (e.g., notifications, buttons, and pop-up windows)
Where	*Places*	Physical environment (e.g., buildings, rooms, office premises, streets, and parks)	Digital environment (e.g., social networking platforms, mobile applications, and video games)
	Events	Co-location (e.g., physical presence at conferences and parties)	Distributed events (e.g., teleconferences, webinars, and a remote presence)
Who	*Agents*	Animate beings (e.g., human beings, pets, and animals)	Artificial beings (e.g., chat-bots and artificial intelligent agents)

Another vein of research and design has sought to measure serendipity despite the highly subjective nature of the phenomena. For instance, Losada et al. [48] attempted to predict the probability of experiencing fortunate discoveries by exploring phone calls and Bluetooth-based social interactions between employees in organization during a nine-month experiment. The authors concluded that analysis of social networks could reveal information about experiencing serendipitous encounters when establishing new connections. Niu and Abbas [49] further proposed a framework to stimulate users' curiosity through modeling surprise and value in recommender systems.

Some recent commercial services have been aimed at increasing serendipitous encounters among people. Examples include the Serendipity Machine (serendipitymachine.com) social platform and the Seats2Meet (seats2meet.com) service. Both are driven by the vision of Society 3.0 [50], which refers to virtual communities that contribute to social capital, business networks, and ecosystems through social networking and cocreation. The primary objective of Seats2Meet is to arrange social interactions through real-time, location-based networking. The service provides software to support and facilitate connections, follow-up activities, and value creation. Unfortunately, so far, the service has been little used, possibly because it is firmly tied to physical meeting points.

These studies and designs have demonstrated that understanding context plays a vital role in increasing the chances for the element of surprise to occur. However, not all unexpected and impromptu connections evolve into beneficial, that is, serendipitous experiences. People also have to recognize opportunities and act promptly. Although the concept of serendipity has inspired various system designs, earlier CSCW and HCI investigations narrowly focused on intraorganizational boundaries and tended to simplify serendipity by treating only the initial phase of the serendipity experience that enables chance. While enabling chance encounters is worthwhile, we call for design endeavors that aim further—towards social serendipity. This end demands new design approaches to turn mere chance and contingency into beneficial outcomes with the help of algorithmic systems. We propose that knowledge work can benefit from designs that better utilize diversity and dissimilarity principles to bridge people with different expertise. Technology can also take more agencies in assisting knowledge workers to follow up on their new connections and uncovering professional values that could otherwise go unseen due to human biases [51]. We continue this line of discussion based on the empirical study, as follows.

3. Methodology

To collect knowledge workers' experiences of serendipitous social encounters, we employed an international, English-language online survey with open-ended questions accompanied by Likert statements. We anticipated that finding relevant respondents with interesting experiences would be challenging. We, therefore, decided to focus on a qualitative approach by including several open-ended questions rather than a quantitative approach with many closed-ended questions that would not allow an in-depth understanding of the experiences.

3.1. Online Survey Structure. To explain the purpose of the survey, we introduced serendipity as the experience of unexpected social encounters that are perceived as fortunate and result in personally valuable interactions, networks, information, and other outcomes from social interactions. This definition was informed by the aspects and essential elements of serendipity drawn from the theories introduced in Section 2. We also supported this explanation by including a storyboard (Figure 1) to the introductory part of the survey.

After reading and signing the informed consent form, the respondents proceeded to the questionnaire, which had five sections (Table 2). The first section focused on the respondents' experiences of social serendipity and included questions about the total number of such encounters and short descriptions of their most memorable ones (Q1 and Q2 in Table 2). The second section asked for a detailed description of their most memorable or important experience (Q3 and Q4). Earlier research has demonstrated significant role of context [35], so the third section included 7-step Likert statements and open-ended questions about the physical, social, and personal contexts of the reported encounter (Q5–Q10). The fourth section asked for details about attitudes, similarities, and differences with the encountered person (Q11–Q14), as well as follow-up activities and valuable outcomes (Q15–Q17). Section five investigated the respondents' opinions and ideas about ICT-enhanced serendipity (Q18 and Q19). Finally, the survey collected the respondents' demographic details, attitudes toward technology, social media use, and general social networking practices.

Table 2: Online survey structure and verbatim questions.

I. General information regarding serendipitous experiences
Q1. Overall, how many serendipitous encounters do you think you have experienced?
Q2. What portion of these was somehow successful and led to positive results in your work?
II. Most memorable experience
Q3. Please shortly describe what kinds of experiences or moments they were. Did they have something in common?
Q4. Now, think of your personally most important experience or moment. If it is hard to decide between a few alternatives, it might be best to concentrate on the one that you remember the best. For the selected experience, we would like you to verbalize the first encounter from your perspective in as much detail as possible. Feel free to write a short story of the moment in your own way.
III. Context
Q5. The place where we met was to me: (1) very unfamiliar–(7) very familiar.
Q6. The other surrounding people around were to me: (1) very unfamiliar–(7) very familiar.
Q7. What were your first feelings or impressions after the meeting?
Q8. Your willingness to socialize at that moment: (1) very unsocial–(7) very social.
Q9. Busyness: (1) I was in a great hurry–(7) I was not busy at all.
Q10. Energy level: (1) I was very tired–(7) I was full of energy.
IV. Attitude, follow-up, and valuable outcome
Q11. Please rate how different your interests were from the person you met: (1) very different–(7) very similar.
Q12. If you identified common interests with the person you met, please briefly describe what they were about.
Q13. Please rate how different your personality is from the person you met: (1) very different–(7) very similar.
Q14. Please describe how comfortable you were while interacting with the person you met: (1) very awkward–(7) very comfortable.
Q15. What were the long-term results of this encounter? How has this encounter been valuable to you and your work?
Q16. How and when did you realize that this encounter or new connection was valuable?
Q17. In your opinion, what made it possible to take advantage of this encounter or new connection?
V. Ideation on ICT-enhanced serendipity
Q18. How do you think information technology (e.g., mobile applications, online services, and social media) could support serendipitous encounters?
Q19. What kind of technical features could help people discover each other or help people meet?

The online survey questions were partly inspired by McCay-Peet and Toms's [23] model of the serendipity process. For instance, the third section was aimed at eliciting details on contextual factors that could serve as triggers for serendipity. Q15–Q17 were intended to collect details on the phases of follow-up, connection, and valuable outcomes, according to the process model.

3.2. Recruitment and Respondents. We targeted knowledge workers, whom we defined in the survey as *"people who develop or use knowledge in their job (e.g., researchers, engineers, consultants, teachers, and coaches)."* The respondents were recruited through mailing lists and flyers at academic events, as well as web-based platforms, such as the Call for Participants service, (callforparticipants.com) Facebook, LinkedIn, and Twitter. As an incentive, the respondents were invited to take part in a raffle for an iPad and two \$50 Amazon vouchers. We closed the survey after collecting 50 responses over several months, which we considered a sufficient amount for qualitative analysis. Of the 50 respondents, 37 provided sufficient details about serendipitous social encounters that had effects on their professional life. The other 13 stories were excluded because they either were focused on nonprofessional benefits such as love stories and friendships or were about general opinions regarding serendipitous experiences without specific examples from their lives.

The respondents' ages ranged widely from 22 to 77 years (M: 35 and Mdn: 30). Nineteen respondents were female and 18 male, and they primarily came from European and Slavic backgrounds: Finnish ($N = 12$), Russian ($N = 11$), British ($N = 3$), and Czech ($N = 3$). The rest were Slovakian, Greek, American, Indian, Kazakh, Persian, Italian, and Yemeni. At the time of completing the survey, the majority of the respondents were pursuing doctoral studies ($N = 15$), five were Master of Science students and one was an undergraduate student. Seven respondents had received undergraduate degrees, six Master of Science degrees, and three doctoral degrees. Researcher ($N = 9$) was the most common occupation among all the respondents. Some reported working in ICT as, for example, data architects, software engineers, and computational physicists. The remaining professions included, for example, business owners, health care experts, designers, management, and sales specialists.

The majority of the respondents indicated that they had experienced only a few serendipitous social encounters in their lives (average: 7, minimum: 1, maximum: 30, and standard deviation: 6). For 21 respondents, most of the new, unexpected connections (>50%) had evolved into serendipitous experiences that positively affected their work. For 13 respondents, less than 50% of the encounters were useful in their professional life, and for the remaining respondents, it was hard to estimate proportions of professionally valuable serendipitous social encounters. Most respondents reported that they were somewhat social and somewhat proficient in becoming acquainted with new people (M: 4.96 and Mdn: 5). On average, the majority ($N = 25$) had 100–500 friends in social networking services, while nine respondents had more than 500 connections and three respondents had fewer than 100. Interestingly, many respondents were not especially active or strategic in networking (M: 4.37 and Mdn: 5). They

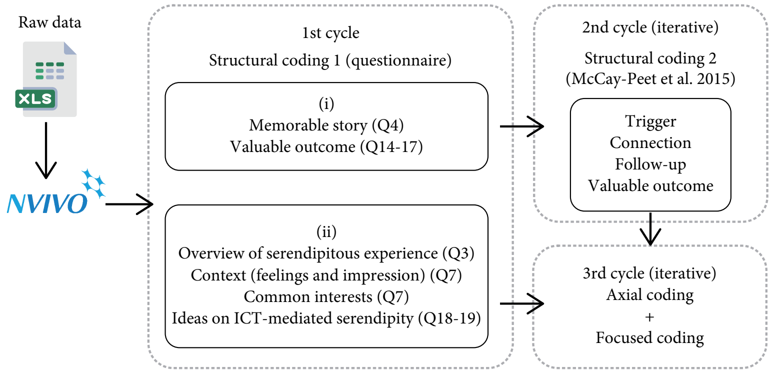


FIGURE 2: The process and methods of qualitative coding.

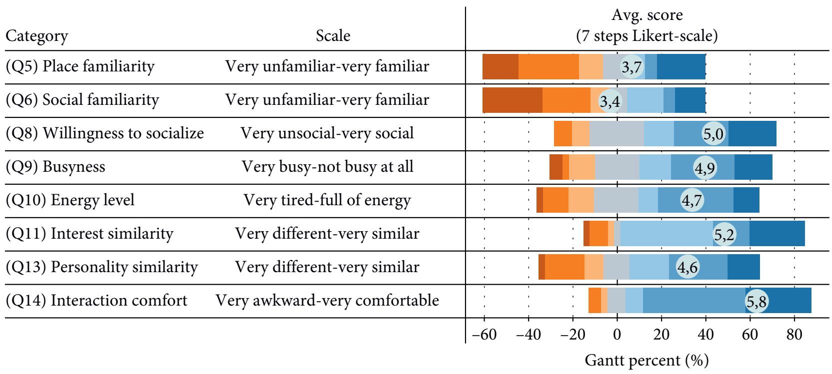


FIGURE 3: Attributes of the respondents' serendipitous social encounters.

considered themselves to be empathetic (*M*: 5.6 and Mdn: 6) and somewhat pleasant and easy to interact with (*M*: 5.3 and Mdn: 5). Most agreed that their work related to knowledge and information acquisition (*M*: 5.73 and Mdn: 6), and their jobs required active collaboration with others (*M*: 5.36 and Mdn: 6). The respondents generally were positively oriented toward technology (*M*: 5.32 and Mdn: 6). Overall, their job descriptions and backgrounds fit well with what could be considered to be knowledge workers. Although not used in the statistical analysis (e.g., comparisons between different groups based on background variables), such data could help planning follow-up studies with more quantitative approaches.

3.3. Data Analysis. The quantitative data was processed in Tableau, (http://www.tableau.com), and the answers to the open-ended questions were imported into NVivo software (qualitative data analysis software, http://www.qsrinternational.com/nvivo). The length of the reported stories varied remarkably from a minimum of 50 words to a maximum of 570, with an average of 246, median of 191, and standard deviation of 148. The coding process consisted of several cycles (Figure 2) including structural, axial, and focused methods [52]. In the first cycle, we applied structural coding that allowed grouping the raw data into top-level categories from the survey questionnaire and resulted in two coding sets. The first set consisted of answers from Q4 and Q14–17 (Table 2) describing the most memorable stories and valuable outcomes recognized by the respondents. The second set comprised a general overview of serendipitous experiences (Q3), feelings and impressions (Q7), common interests (Q12), and ideas about ICT-enhanced serendipity (Q18-19). Next, for the first set, we conducted another round of structural coding based on the serendipitous experience process in McCay-Peet and Toms's [23] model (trigger, connection, follow-up, and valuable outcome). Then, we utilized axial coding, which included line-by-line analysis and deconstruction of the data into emerging categories for each phase of the process. Finally, focused coding was applied to identify the most frequent categories. The coding cycles were iterative and involved the three authors, who conducted the coding separately and cross-checked each other's categorizations to validate the quality of the coding structures.

[...] It had happened around two years ago. I was at the ski slope trying to do downhill for my second time. It was a tough day, and I was completely exhausted from falling, hitting other people and trees [...] A person, who was passing by few times, stopped and started to teach me how to turn and brake. He spent around 20 minutes with me, and I was already able to do something. [...] In the evening, I was very excited about this story. During the next day, [...] I have met him again but on the lift. We had a nice life deep conversation and spent some time skiing. During our short lift-talk, we ended up realizing that our research interests and fields are almost the same. We have changed cards and now have a collaboration going! [...] (R1, Russian male, 25 y.o., Researcher, length cut 42%)

- Contextual triggers: public place
- Interaction triggers: unsought social interaction, communication, repeated encounter, got assistance or advice
- Connection: revealed commonalities
- Follow-up: contacts exchange
- Valuable outcome: primary professional, secondary personal

Figure 4: An excerpt from a respondent's story of a chance encounter during leisure activities that led to professional collaboration.

Being part of three [anonymized organization] projects, I spent quite much time on their premises. [...] So kitchen was a meeting point for everyone. [...] A place, is pretty open and people are talkative and quite friendly. I kept bumping into a person, and every time we were just having short chats and talks. So, one day as I was waiting for my tea, I bumped into my (now) colleague, and he started to tell me in detail what was his startup about. One thing led to another, and we decided to set a proper meeting where we would sit down and discuss if I could actually help him or not. It turned out that he needed someone who could help him with design in his current project and I offered to do that. Since then, we're still working together on this project, and so far we're having good results and nice time together. [...] (R4, Greek female, 41 y.o., Graphic/UI Designer, length cut 52%)

- Contextual triggers: work premises
- Interaction triggers: communication, repeated encounter
- Connection: revealed complementarity
- Follow-up: repeated social interaction
- Valuable outcome: primary professional

Figure 5: An excerpt from a respondent's story about how the work environment facilitated a serendipitous social encounter.

4. Findings

In this section, we provide an overview of the serendipitous social encounters, including quantitative and qualitative observations and examples in excerpts from the respondents' narrations. We next describe each phase of the serendipitous experiences, following the process model by McCay-Peet and Toms [23] but contextualizing it to the social aspect of unsought encounters.

4.1. Overview of the Serendipitous Encounters. The quantitative results in Figure 3 demonstrated that the majority of respondents experienced serendipitous social encounters mostly in unfamiliar environments during both physical and social contexts. In general, they were energized and positively oriented to socialize and had free time to take advantage of opportunities. Regarding personal attraction, the respondents found these interactions with other persons to be very comfortable, and they were able to identify similarities in their personalities and interests.

From an experiential viewpoint, the encounters led to various positive feelings such as excitement, thrills, curiosity, interest, and a sense of belonging and importance. As expected, due to the nature of serendipity, the majority of the respondents explicitly mentioned that they felt lucky and could not believe in such coincidences. They also interpreted their experiences as uncontrollable and unsought. Some assumed that fate or providence had enabled these encounters, while others considered them to be the results of personal traits such as openness, willingness to communicate, and the ability to step outside their comfort zone. The following example depicts how a respondent recognized both aspects—the roles of luck and individual personality traits:

> *"The fact that we were both talkative and we found each other quite interesting people made this possible. Besides a little bit of luck, because he was looking for someone with my set of skills. At that time, I was also very active, and it came quite naturally to me to grasp this opportunity and work with something that I found both challenging and interesting." (R4, Greek female, 41 y.o., Graphic/UI Designer)*

The reported stories represented a high diversity of contextual characteristics and illustrated the entire social serendipity process, from the moment of a chance to recognize and establish value in one's professional life. The analysis of the qualitative data showed that the temporal distance between the first encounter and the positive outcomes varied from days to decades. The majority of the stories were very detailed, often with descriptions of preconditions such as personal feelings and chains of actions that led to the encounters and obtaining valuable outcomes.

Social media is accelerating the process. [...] The most recent: last year, a man connected to me on Facebook after reading my comments to a mutual friend. He just traveled to place, where I live, to give a workshop and wrote to ask whether I had time to meet him two days ago. We met and had coffee. We have some common experiences, values and interested in sustainable development, coaching, ethics. We found an immediate prospect to work together. He also introduced me to another person whom I met last night at the event, and she also wants to work together. (R37, British female, 77 y.o., Head of training & development, length cut 43%)

- Contextual triggers: ICT-mediated environment
- Interaction triggers: communication
- Connection: revealed commonalities
- Follow-up: repeated social interaction
- Valuable outcome: primary professional, secondary social

Figure 6: An excerpt from a respondent's example of an ICT-mediated environment that increased the probability of social serendipity.

Table 3: Categories of contextual and interaction triggers identified in the respondents' stories.

Trigger type	Category	Examples from the stories
Contextual	Public place (10 stories)	Streets, airports, clinics, train stations, hostels, bus stops, saunas, cafés, and ski slopes
	Professional event (7 stories)	Conferences, exhibitions, and seminars
	ICT-mediated environment (7 stories)	Phone and video calls, emails, and social network services (e.g., Facebook)
	Work premises (6 stories)	Kitchens, meeting rooms, offices, and halls
	Education premises (5 stories)	Campus areas and group classes
	Social event (2 stories)	Volunteering events and fairs
Interaction	Direct communication (29 stories)	Introducing each other, small talk, useful information exchange, official meetings, and discussions of fields of interest, problems, and opportunities
	Repeated encounters (9 stories)	Bumping into the same person several times, attracting attention and motivating follow-up
	Unsought interactions (9 stories)	Introduction of two actors by a familiar person (e.g., a friend or supervisor) when they least expect it, initiation of interactions by stranger
	Receiving assistance or advice (8 stories)	Assumption of the role of mentor by the recently encountered person, leading to professional follow-up or valuable outcomes

To provide a detailed picture of the collected experiences and to illustrate the coding process of the qualitative data, we selected three stories that well represented the diversity of the gathered data. The first case (Figure 4) referred to a chance encounter that was initiated during leisure activities and evolved into a serendipitous experience with professional collaboration as its valuable outcome. The informal contextual setting, combined with the respondent's problem, triggered interactions with a stranger, who not only helped in the current situation but also became a collaborator after later social interactions.

Another story demonstrated how a creative working environment could serve as a networking space and increase the probability of serendipity (Figure 5). Here, repeated encounters provided the element of surprise that motivated the actors to interact and revealed previously hidden opportunities for cooperation. The final example (Figure 6) illustrated how technology-mediated interactions in social networking services could lead to face-to-face meetings and result in the planning of professional partnership. In this context, the respondent's follow-up of the opportunity and willingness to interact played significant roles in turning chance into serendipity.

4.2. Analysis of the Social Serendipity Process. McCay-Peet and Toms's [23] model was used as a framework to identify different phases in the narratives (trigger, connection, follow-up, and valuable outcome), which determined the structure of this subsection. We created a bottom-up categorization of the different instances of each phase in the stories, supported by the examples from reported stories.

4.2.1. Triggers. In the serendipity process, a trigger constitutes a moment when a chance encounter happens. In other words, it refers to the general context, social situation, or setting that affects an encounter. Earlier research classified triggers based on sensory qualities [23] or types of objects and environments [14] due to the primary focus on unexpected encounters with information. We investigated serendipitous encounters from the perspective of making new valuable connections, so we distinguished between two types of triggers: contextual and interaction triggers. Contextual triggers refer to environmental settings that spark perceptions of serendipitous experiences, while interaction triggers cause unsought interactions with the person encountered.

Table 4: Categories of the connection phase identified in the respondents' stories.

Category	Examples from the stories
Revealed commonalities (33 stories)	Job fields, professional and personal interests, hobbies, life goals, work styles, shared social ties, education fields, and life experiences
Revealed complementarities (17 stories)	Complementary professional skills and needs, different backgrounds and viewpoints

In this study, the most common type of contextual triggers was related to public places (Table 3). In addition to the example in Figure 4, the respondents reported meeting interesting others on public transportation and in train stations, restaurants, hotels, and even the streets. Other common contexts included professional events such as conferences, seminars, and professional exhibitions.

Some of the stories featured ICT-mediated environments (Figure 6) such as video conferences and Internet sites (discussion forums and websites), and the use of social media services could indeed increase the probability of unsought encounters. Many mentioned that they believed that ICT could simplify the process of finding relevant others and increase awareness of potential collaborators:

> *"ICT can help connecting professionals online based on, e.g., interests, professional targets, or ambitions and offer "blind date" type of connection possibilities, a little bit like "people you may know" on LinkedIn, but "people you have common interests in." I cannot really visualize the technical features other than by saying that some algorithm which can provide such "blind dates," and then based on the users' feedback learns to understand better what kind of people would more likely be useful, new encounters." (R21, Finnish male, 48 y.o., Sales Director)*

One respondent went so as far as to envision computational platforms that could enable the advanced formation of genuinely global virtual collaboration communities across any demographic boundaries:

> *"ICT solutions might collect data about my and others' daily digital exhaust, and algorithmically finding new potential connections for me to explore. I envision, informal spaces for team-based work. I'm not talking about co-workings or lofts, but about a platform (social or virtual) where a person leading some project will be able to find and recruit, in a way, required specialists and communicate with them (no matter what language they speak or where they live)." (R29, Finnish male, 39 y.o., Researcher)*

Despite the generally positive attitudes toward technology-mediated serendipity, some respondents seemed to regard existing solutions as having limited ability to facilitate follow-up activities and value creation among people:

> *"ICT can increase so-called coincidensity—the increase of chance encounters between a diversity of people—but in many cases, it does not support the value creation, which is the critical element in serendipity." (R32, Finnish male, 62 y.o., Founder and partner)*

Additionally, the respondents were concerned about the trustworthiness of computer-mediated communication and the reliability of digital persona. Such attitude could be explained by the gap between trust in online and offline (face-to-face) interactions and the variance in this respect in different digital communication media (e.g., chats vs. video calls), as implied in the following example:

> *"I believe there should be a way of having social media profiles validated as accurate and trustworthy. I do trust people, but I would feel better when accepting connection requests from strangers if there was a way of checking their profile is true, similar to the blue tick that celebrities have on Twitter. [...]" (R35, British male, 48 y.o., Data architect)*

In addition to contextual factors, social serendipity experiences could be initiated through unsought interactions with the encountered person, in words, through interaction triggers (Table 3). For instance, repeated encounters with the same person could easily catch one's attention (i.e., the familiar stranger) and motivate initiating interactions, possibly leading to potentially valuable outcomes. In other cases, unsought social interactions occurred when an individual was least expecting to socialize with others, increasing the perception of serendipity. The majority of the stories reported that small chats and informal discussions triggered serendipitous experience: ongoing conversations gradually led to identifying mutual interests and synergies.

Another type of interaction trigger was unsought advice or assistance. In addition to the stories represented in Figures 4–6, a representative example of interaction triggers came from the following story:

> *"[...] My department was taking up an international project with a foreign university. Professor asked me if I could [...] welcome the guests [...] and look after their comfort and other arrangements. Since I held him in high regard, I could not refuse. [...] When I first met them, I was hesitant, but they were very friendly. [...] We discussed our research areas and interests, and those matched well. By the time we reached the Guest House, they had developed a liking for me! Later that year, I came to know of a Ph.D. award and was encouraged by my supervisor to apply for it. [...] The first and most crucial step was to identify and approach a prospective supervisor from another country. [...] I had forgotten all about the two visitors that I assisted, but my supervisor happened to recall my pleasant meeting with them and suggested to contact them. [...] I applied for the award, got selected, and now [...] working with these two lovely and intelligent professors on my Ph.D. topic." (R6, Indian female, 26 y.o., Doctoral student, story length cut 42%)*

Table 5: Categories of the follow-up phase identified in the respondents' stories.

Category	Examples from the stories
Repeated interactions (30 stories)	Multiple meetings or chats that gradually lead to identification of potentially valuable outcomes
Contact information exchanges (6 stories)	Capturing opportunities for later, for instance, by exchanging business cards or contact details

In other words, the respondent was compelled to interact with foreign visitors and did not expect that these connections would be useful one day. In addition, the supervisor's advice triggered the connection phase of the serendipitous experience, leading to follow-up with foreign professors and achievement of a valuable outcome.

Overall, the data revealed that chances for serendipitous social encounters could appear not only in unknown environments but also in familiar places. Although it was hardly surprising that knowledge workers met at events, it was interesting to note that the respondents also reported unsought and unexpected experiences in contexts where people were generally in socially open moods and mentally prepared for new opportunities. In addition to contextual factors, the experience of social serendipity could be initiated through interactions with other individuals and surroundings. The stories usually included multiple triggers; for example, both contextual and interaction triggers led to the next phases of the experience: connection and follow-up.

4.2.2. Connection and Follow-up. In the traditional notion of serendipity, connection refers to the phase when a person reflects on the trigger based on their background and knowledge, leading to identification of possibly valuable outcomes. In this study, connection refers to proactive, reciprocal reflection on triggers and the interactive process of obtaining an understanding of how to benefit from serendipitous social encounters. Such sense-making of the encounter is built on recognizing common ground as people get to know each other and exchange information, often during their first meeting (Table 4).

The respondents described various commonalities between themselves and their recently met persons. While the majority of the respondents referred to having similar jobs and interests in general, some specifically mentioned shared research interests. There were also some examples of identification of similar personality traits, hobbies, and even life experiences. One story brought up the similarities and differences in of a good match in which the heterogeneity of the background and viewpoints and the similarity of goals and interests had benefits for both professional and personal contexts:

> *"I took part in a video lecture, sitting behind a Tv-screen in a meeting room. [...] The discussions were difficult due to a bad connection, but I remember how excited I was when someone started to comment about a book related to the lecture. I had just finished it, and it had also been important for my thinking. [...] We later met face-to-face with that person, and since that day [...], we have been helping each other's growth both in personal as well as in professional life. [...] We both work on dissertations, and although they sound to be of subjects quite far from each other, we are interested in the same phenomenon, but from a different point-of-view. We also share ideological goals and interests in personal lives. [...] With quite a different background and a different logic/mindset, we have grown to the benefit of our differences and to learn from each other instead of "only" sharing similar experiences and thoughts." (R34, Finnish female, 42 y.o., Researcher, story length cut 52%)*

The connection phase was crucial to obtaining an understanding of how to benefit from a serendipitous social encounter, whether to solve urgent problems or access new opportunities or directions. However, to achieve the optimal benefit and turn pure chance into serendipity, one had to invest in the follow-up phase. In social serendipity, follow-up refers to the deliberate social interactions with the encountered person needed to obtain a valuable outcome. These interactions can include follow-up messaging, meetings, and more active collaboration. Thus, the connection, together with the follow-up phase, can be said to distinguish serendipity from pure luck or chance. One respondent deeply contemplated this fact:

> *"In all my serendipitous encounters, there are two common aspects—I have always been expecting nothing (no follow-up) after the first contact and so was surprised that actually, this particular case developed somehow. As a second, I always had to invest some professional and personal efforts to explore the relation and get some long-term results. [...]" (R16, Czech male, 33 y.o., Researcher, story length cut 78%)*

Our analysis identified two main categories of follow-up actions and interactions (Table 5). The first one referred to later capturing opportunities when an individual realized the possible value in the initial encounter but could not yet identify how or where to apply this potential. In our data, such actions included, for instance, the exchange of contact information. The majority of the respondents, however, focused their narrations on the process of identifying mutual benefits during repeated social interactions such as meetings and phone calls. It is worth noting that many stories had a chain of follow-up actions, possibly including a delay between the connection and follow-up actions. Often, much information needed to be exchanged before a mutually beneficial topic of collaboration was discovered.

While expecting technology to be a powerful tool for connection and follow-up interactions with recently established ties, some respondents noted that the ways in which social media platforms were designed might limit their use in the process of strengthening relationships:

Table 6: Categories of valuable outcomes identified in respondents' stories.

Category	Examples from the stories
Primary—professional (37 stories)	Professional collaboration, vocational growth, business partnership, and career planning
Secondary—social and personal (16 stories)	Friendship, extended social networks, mentoring, useful information retrieval, self-improvement, graduation, and educational awards

"I would like to know more about my network, but at the moment it is very superficial, I see only a page with a few works and a list of their education and experience. It tells me little about their personality, their hopes and expectations, their goals and aspirations. As a consequence, I have many contacts but not so many friends or people I know personally. I think faster networks and virtual reality meetings will help to create lasting impressions and build stronger relationships. I regret that I can only text message to my LinkedIn contacts when a short virtual face-to-face meeting would be much nicer and more humane." (R35, British male, 48 y.o., Data architect)

4.2.3. Valuable Outcomes. Valuable outcomes refer to the positive outcomes of serendipitous social encounters. All the analyzed stories resulted in professionally valuable outcomes, and some were also followed by additional personal and social benefits (Table 6). The valuable professional outcomes included examples of collaboration and partnering, gaining employed, changing jobs, and deciding on career directions. In fact, in some cases, an encountered person became not only a work partner but also a good friend in personal life. Some respondents explained that the encounters helped them build and extend their social networks, while others cited personal achievements as benefits. For instance, when assisted with useful information, one might gain support to finalize one's education or obtain educational rewards. Finally, the respondents also mentioned self-improvement, referring to learning new perspectives, getting additional training, solving personal problems, improving professional skills, and changing one's worldview and behavior for better. We consider this spectrum of beneficial outcomes to consolidate the idea of serendipity as a fruitful experience to universally seek in professional life.

The respondents mentioned that not only chance or luck led to beneficial outcomes. Other relevant aspects of serendipity included, for instance, openness to new opportunities, curiosity, personal experiences, and professional skills, which allowed taking advantage of serendipitous encounters:

"Many chance encounters do not lead onto anything significant beyond a pleasant encounter, professionally or personally, but occasionally a relationship can lead to a meaningful and important event or significant change. [...] For instance, when a chance encounter in a corridor in a building led me to become involved in a predictive analytics project for a business partner. Such events do not happen often, but they happen with sufficient regularity that I am minded to network with people and be open to new possibilities [...]" (R35, British male, 48 y.o., Data architect, story length cut 61%)

Individual qualities were essential to realize the opportunity for a fortunate encounter and benefit from serendipity. For example, openness was needed to have the appropriate mindset and experience to perceive the opportune moment. In other words, impact took place when there was a space for it in one's life.

5. Discussion

The results imply that knowledge workers' experiences of serendipitous social interactions manifest in various situations beyond organizational boundaries and offer a wide range of perceived benefits. The data also demonstrate the diversity of the contexts of the encounters and the triggers that affect the matching process. The subjectivity of drawing a line between serendipity and mere luck results in variance in the intensity of the experiences and effects the respondents included in their stories. With our survey design and recruitment efforts, we managed to gather relatively strong experiences of social serendipity. However, we expect that we were unable to collect other more mundane instances of these experiences. In the following, we highlight a few of our qualitative findings that we consider to best characterize the concept of social serendipity. In addition, we discuss the seemingly minor role of technology in social serendipity and outline design directions to enhance serendipitous experiences with technology. Finally, we address the methodological limitations of the study.

5.1. Key Characteristics of Social Serendipity. The first characteristic addressed is that social serendipity can be seen as an instance of aesthetic experience [53]. To be more specific, serendipity is a well-formed, complete experience that is subjectively unified, recognized, and interpreted. It has a beginning and an end and unfolds over time, producing a satisfying emotional quality—the joy of unsought fortune—that permeates the entire process. The flow of the serendipitous experience thus is characterized by the purposeful connection of all phases coupled with the subject's actions and interactions with the physical and social environments. In the following, we reflect on each phase of social serendipity identified in the stories and contrast them to earlier research on the conceptualization of serendipity.

Theorizing serendipity, researchers [14, 23] revealed types of triggers that make sense in the context of information retrieval in which the experienced outcome generally relates to knowledge discoveries. In contrast, social serendipity consists of building valuable interpersonal relationships and thus is primarily driven by interactions between people. Our findings, therefore, distinguish between contextual and interaction triggers. The stories indicate that experience is rarely sparked by only one type of a trigger but instead is more commonly caused by a complex combination of several triggers. While contextual triggers create a fruitful scene for serendipity to occur, interaction triggers play a more significant role in the experiences of serendipitous encounters because they lead to the next phases.

In a traditional notion of information serendipity, the connection phase is characterized by unconscious processing of the triggers and content, but our findings on social serendipity demonstrate the qualities of proactive and reciprocal sense-making. The connection phase in interpersonal interactions centers on two primary approaches: through direct social interactions, people either reflect on commonalities between each other (e.g., to build trust) or look for complementary characteristics (e.g., to identify reasons to collaborate). The stories present two types of follow-up actions and interactions: (i) repeated social interactions to strengthen connections and explore mutual benefits; and (ii) exchanges of contact information to capture opportunities for later use. To turn pure luck and chance into social serendipity, both actors have to invest effort and commitment during the connection and follow-up phases.

As mentioned, the outcomes in the traditional notion of serendipity generally are related to knowledge discovery. Of course, in social serendipity, knowledge discovery is also one of the benefits complementing the primary value: new beneficial relationships with other people. In fact, making meaningful connections with other people contributes to information flow and knowledge creation. All the examples from our survey aimed at collecting stories about serendipity in professional life explicitly refer to work-related benefits (e.g., finding new jobs and collaborators) as the major valuable outcomes. However, many stories indicate that professional gains are often accompanied by secondary benefits (e.g., friendship, mentoring, and self-improvement), which enable even stronger social bonding. Moreover, recognizing that one is experiencing social serendipity and the related satisfying emotional qualities can be perceived as beneficial outcomes of the experience.

5.2. The Role of Technology in Social Serendipity. As addressed in Section 2, earlier HCI research demonstrated considerable interest in developing ICT to support social matching and chance encounters. Although only a few narrations highlight the role of technology in contributing to these experiences, the respondents believed that ICT can help identify opportunities for serendipitous professional collaboration and assist in matching and communication processes. While this study does not allow generalizing the prevalence of ICT in knowledge workers' experiences of social serendipity, the minor role of technology in this sample highlights the importance of reconsidering what kinds of computational approaches can more effectively foster serendipitous social encounters. We claim that the presented empirical findings on the social serendipity process can support the definition of meaningful design goals and directions for future ICT. We especially apply our thinking to social matching services [54] and people recommender systems [55] because they both represent a more proactive paradigm of technology. In addition to presenting relevant information, such technologies make complex inferences and use algorithmic logic to suggest alternatives to users and actively help their decision-making.

Current matching services and people recommender systems often employ algorithms based on two well-established social networking mechanisms. First, matching is primarily driven by the homophily hypothesis [56], which states humans tend to connect mostly with those similar to themselves. Second, they follow the triadic closure hypothesis, which holds that new connections are likely to form between friends-of-friends [57], that is, between actors who already have strong, trustworthy ties [58]. The human preference for like-minded others and trust in new connections introduced by familiar others are even visible across some stories collected in this study. However, these principles might be detrimental to knowledge work because they contribute to the formation of social groups that consolidate viewpoints, thus reducing the creation and distribution of new knowledge and opinions. This phenomenon is referred to as an echo chamber [59], and social media services have recently received criticism for supporting echo chambers [60].

Responding to these challenges, researchers in the domain of recommender systems have started to promote serendipity as a feature that design should support [20]. The primary objective of such serendipity-enhancing systems is to increase the quality of recommendations by diversifying content and enabling novelty and unexpectedness. Diversification of the recommended content can increase the probability of serendipity but does not always lead users to perceive it. After all, intentionally using digital artifacts to identify new potential connections inherently reduces the element of unexpectedness. de Melo [14] claimed that to provoke serendipitous experiences, systems have to implement all three qualities—new discoveries, unpredictability, and value. While the first two qualities can be achieved through approaches such as randomization [61, 62] and defamiliarization [63, 64], facilitating the creation of value for the user is the most challenging task.

Inspired by de Melo's notion of serendipitous systems [14], we call for digital services that can contribute to the entire processes of social serendipity. We propose a user-centric idea of ICT-enhanced social serendipity experiences with three vital process-based elements: (i) an initial phase of introducing opportunities for professional networking and interpreting their inferred relevance, (ii) encouraging direct social interactions for value creation, and (iii) maintaining social awareness of established connections through facilitated follow-up processes. In the following, we discuss these elements in relation to potential computational service qualities. We primarily focus on aspects relevant to the HCI community.

5.2.1. Introducing Opportunities and Facilitating the Interpretation of Inferred Relevance. Our findings imply that experiences of social serendipity are often initiated in unsought, face-to-face chats and discussions. Optimistic expectations of how computer-mediated communication might lead to serendipity were rare among the respondents, and most thought that the first interactions should take place face-to-face to establish social bonding and trust. These findings could be explained by the fact that many of the respondents were middle-aged or older citizens possibly less trustful and open to the use of technology than millennials [65].

Indeed, in virtual environments, many issues tend to limit interactions with unfamiliar people. For instance, privacy restrictions might inhibit making new connections. The lower psychological pleasure from computer-mediated communication than face-to-face interactions might reduce motivation to chat with strangers. The lack of exploratory, open-minded social settings in virtual environments might hinder people from investing sufficient time for exploring networking opportunities and mutual interests. In the early phases of the social serendipity process, the probability of experiencing the unexpected depends on the content such as other users' profiles and their presentation in the user interface. In people recommender systems, the degree of how surprising and seemingly relevant the suggested new contacts are can affect users' acceptance and willingness to interact with them.

The current standard of introducing people to each other in social network platforms (e.g., recommendations of whom to connect to or follow) is based on a simplistic, list-based approach using profile pictures, names, and brief biographical summaries. This way of delivering recommendations can be enough for lightweight, low-risk decisions, such as whom to follow or to identify existing acquaintances on a platform. However, due to the various needs of partnering and collaboration (e.g., from seeking a mentor for personal vocational growth to building a community), finding valuable, meaningful people in knowledge work requires enhanced content and profiling and advanced visualization techniques to deliver recommendations. In this regard, the variety of data about people available in social media services could help build more comprehensive user profiles and analyze existing ties in the larger social network [66]. Social network analysis could also enable computational identification of potentially relevant new and weak ties. On a broad level, interactive, exploration-supporting visualization of several perspectives of potential collaborators in a social network graph could facilitate the experience of surprise and luck. Positioning a user in a social graph and explaining the inferred relevance with recommended people (e.g., by visualizing the similarities and differences mentioned by R34) thus could trigger serendipity and lead to the connection phase.

The connection phase is primarily a mental process of analyzing information about the new encounters and reflecting on and interpreting the potential value of prospective interactions. To support this phase in ICT-mediated environments, it is essential to create a smooth transition to social interactions, first making users familiar with each other through enhanced profiling. This calls for identifying the types of content that, first, best suit the diverse contexts of professional matching and that, second, knowledge workers consider to be sufficiently valuable to start interacting with an unfamiliar match. In this case, user profiles should present content that steers readers' attention to users' qualities that are timely and relevant to understand, thereby initiating social interactions. Self-created content could be complemented with computational analysis of personal social data and extended with insights inferred from different digital platforms and social media services. Enriching each user's image could be useful to identify unanticipated opportunities, inherent personal qualities, and contextual oddities [67], attributes that might not appear in user-generated content.

5.2.2. Encouragement and Persuasiveness for Value Creation and Maintenance of Social Awareness. The findings imply that even in face-to-face interactions, encounters that result in serendipity could benefit from the *tertius iungens* orientation [68], or facilitators to encourage, initiate, and coordinate interactions with new people and motivators to follow up contacts. In addition to recommending seemingly relevant new connections (i.e., increasing chances), ICT could also facilitate the connection phase and aid identification of potentially beneficial outcomes. While designing for various encouraging features (i.e., persuasiveness) in ICT applications is tempting, achieving an acceptable yet useful degree of persuasion is highly challenging. Current people recommender systems in social media services can be considered to have a reasonably reliable ability to suggest a variety of opportunities, but they have no role in converting these recommendations into actual behavior.

Designing ways to motivate users to perform the necessary actions on the path to realize possible social serendipity calls for audacious design exploration. We propose that this role could be best realized through cooperative synergy between a data-driven analytics system focused on computational analysis and a user capable of making personally relevant interpretations and focusing on the qualitative aspects of social matching. Thus, ICT-enhanced social serendipity could be aimed at achieving human-in-the-loop analytics [69] and augmented intelligence [70, 71] rather than artificial intelligence, that is, supplementing human intelligence with computational capabilities through new forms of HCI. The innate differences between human decision-making and computational logic necessitate careful design of the interplay of these two types of intelligence [72]. For example, considering the well-known paradox of choice [73], instead of increasing the numbers of matching opportunities, perhaps ICT applications should reduce choice anxiety by radically limiting the options of suitable individuals or giving users efficient ways to do so reflectively and systematically.

For example, ICT could provide cues regarding its inferences of how a given social recommendation is relevant to the target user and deliver so-called tickets to talk to initiate discussions and maintain already established ties. Furthermore, as the findings demonstrate, valuable outcomes are generally achieved after long-term follow-up activities, sometimes over the course of several decades. To maintain interpersonal relationships and to support value creation among users, one suitable approach could be to keep users updated on their recommended contacts' recent activities and to inform users of suitable opportunities for face-to-face and virtual meetings. Supporting users in understanding the value of their new and existing connections could be achieved by utilizing advanced personalization approaches. The system could adjust the appearance of users' profiles and contents, gradually revealing more facts about them (enabling the element of surprise) and potentially encouraging direct communication and follow-up avenues. Moreover, by collecting users' feedback on new adaptive profile representations, matching algorithms could learn individuals' preferences and criteria for subjectively perceived relevance. This personalization approach could imitate the repeated encounter triggers revealed in the sample data and thus enhance people's decision-making process in the choice of better matching opportunities for value creation.

5.3. Methodological Limitations. All methodologies impose different limitations on the reliability and validity of findings. In this study, for example, an online survey might not have been considered to be ideal for qualitative research. The sampling could be challenging to control, compromising generalizability, and the truthfulness of the answers could not be verified as in face-to-face interviews. However, the choice to run an online survey with open-ended questions was considered to be the most suitable compromise between the research objectives and practicality. On the one hand, the research gap and our questions called for qualitative accounts to deeply describe and understand the subjective experiences and reported social encounters. On the other hand, the prevalence of knowledge workers' memorable serendipitous social encounters was considered to be relatively low, making recruitment for an interview-based study highly challenging on a practical level. Furthermore, as the respondents self-reported past experiences, recall bias very likely influenced the truthfulness of the reported experiences. This, however, was inevitable because serendipity, in fact, required time to progress and so could hardly be studied as it unfolded.

Despite extensive recruiting efforts, attracting respondents proved to be challenging, which could be seen to support our assumption of the rarity of such experiences. Fortunately, the analyzed responses were very detailed and represented relatively strong, vivid examples of serendipity, which allowed in-depth accounts of the experiences. We speculated that the difficulties gathering data resulted from two main factors. First, some survey visitors might not have understood the concept of serendipity or correlated it with their own experiences. The term has been relatively little used outside academia, especially compared to other related concepts. We anticipated this issue upfront, so the survey introduction also used terms like "chance" and "unexpected encounters that lead to valuable outcome." The welcome page of the survey explained what we sought in layperson's terms. Second, the individual variance in perceiving unexpected [74] might have been another factor preventing people to recognize serendipity. The concept of serendipity was related to chance and luck, so some people may have wrongly assumed that a lack of control was a fundamental element in such experiences. For instance, people with a high locus of control [75] might not have attributed such experiences to chance but, instead, considered beneficial encounters to have resulted from their own efforts and decisions. Little research has addressed cultural differences on the experience of luck and unexpectedness, so we could not assess how well this sample with a specific cultural bias represented the entire human population. Nevertheless, as this study was not aimed at producing generalizable quantitative information (e.g., the prevalence of the phenomenon), we believed that the downsides were tolerable given the seemingly rare research on experiences of social serendipity in knowledge work.

6. Conclusions

Serendipity has been actively theorized in the context of information retrieval, but social serendipity has not been extensively studied with empirical approaches. We conducted an international, qualitative online survey and gathered 37 responses with rich examples of serendipitous social encounters. Our analysis reveals various insights into the nature of social serendipity in the context of knowledge work. In addition to reporting a wide range of aspects and subjective experiences relevant to social serendipity, we extend the theoretical understanding of the serendipity process with aspects particularly applicable to serendipity in social contexts. The gathered insights can be employed to inform the design of more effective and appropriate ICT. While luck and unexpectedness characterize any chance encounters, social serendipity also requires active follow-up to identify their potentially valuable outcomes and active collaboration to realize them. Social serendipity benefits from repeated encounters, social facilitation, and identification of shared goals and reciprocal benefits. Interestingly, however, only a few of the respondents reported that ICT served in roles that somehow contributed to the emergence or support of serendipitous encounters. We, therefore, highlight the question of the meaningful roles of ICT in computationally enabled social serendipity experiences.

Data Availability

The data generated or analyzed during this study are included in this article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was funded by the Business Finland Project Big Match (3166/31/2017 and 3074/31/2017). The authors thank all members of the Big Match project and the participants of the study.

References

[1] L. A. Goodman, "Notes on the etymology of serendipity and some related philogical observations," *Modern Language Notes*, vol. 76, no. 5, pp. 454–457, 1961.

[2] R. K. Merton and E. Barber, *The Travels and Adventures of Serendipity: A Study in Sociological Semantics and the Sociology of Science*, Princeton University Press, Princeton, NJ, USA, 2006.

[3] D. Liestman, "Chance in the midst of design: approaches to library research serendipity," *RQ*, vol. 31, no. 4, pp. 524–532, 1992.

[4] A. Ferguson, "The lost land of serendip," *Forbes*, vol. 164, no. 8, pp. 193-194, 1999.

[5] T. D. Anderson, "Beyond eureka moments: supporting the invisible work of creativity and innovation," *Information Research: An International Electronic Journal*, vol. 16, no. 1, pp. 1–24, 2011.

[6] S. Johnson, *Where Good Ideas Come from: The Seven Patterns of Innovation*, Penguin, London, UK, 2011.

[7] N. K. Agarwal, "Towards a definition of serendipity in information behaviour," *Information Research: An International Electronic Journal*, vol. 20, no. 3, pp. 1–16, 2015.

[8] D. Bawden, "Encountering on the road to serendip? Browsing in new information environments," in *Innovations in Information Retrieval: Perspectives for Theory and Practice*, pp. 1–22, Facet Publishing, London, UK, 2011.

[9] I. Buchem, "Serendipitous learning: recognizing and fostering the potential of microblogging," *Form@ re-Open Journal per la formazione in Rete*, vol. 11, no. 74, pp. 7–16, 2011.

[10] J. E. Nutefall and P. M. Ryder, "The serendipitous research process," *The Journal of Academic Librarianship*, vol. 36, no. 3, pp. 228–234, 2010.

[11] S. E. Alcock, "The stratigraphy of serendipity," in *Serendipity: Fortune and the Prepared Mind*, M. de Rond and I. Morley, Eds., pp. 11–26, Cambridge University Press, Cambridge, UK, 2010.

[12] J. E. H. Bright, R. G. L. Pryor, and L. Harpham, "The role of chance events in career decision making," *Journal of Vocational Behavior*, vol. 66, no. 3, pp. 561–576, 2005.

[13] V. L. Rubin, J. Burkell, and A. Quan-Haase, "Facets of serendipity in everyday chance encounters: a grounded theory approach to blog analysis," *Information Research*, vol. 16, no. 3, p. 27, 2011.

[14] R. M. C. de Melo, *On serendipity in the digital medium: towards a framework for valuable unpredictability in interaction design*, Ph.D. thesis, Faculdade de Belas Artes, Universidade do Porto, Porto, Portugal, 2018.

[15] S. Kiesler and J. N. Cummings, "What do we know about proximity and distance in work groups?" in *Distributed Work*, pp. 57–80, MIT Press, Cambridge, MA, USA, 2002.

[16] D. Vyas, A. Dix, and G. C. van der Veer, "Reflections and encounters: exploring awareness in an academic environment," *Computer Supported Cooperative Work (CSCW)*, vol. 24, no. 4, pp. 277–317, 2015.

[17] P. Jeffrey and A. McGrath, "Sharing serendipity in the workplace," in *Proceedings of the Third International Conference on Collaborative Virtual Environments*, pp. 173–179, ACM, San Francisco, USA, September 2000.

[18] I. Leo, M. De Gemmis, P. Lops, G. Semeraro, M. Filannino, and P. Molino, "Introducing serendipity in a content-based recommender system," in *Proceedings of the Eighth International Conference on Hybrid Intelligent Systems (HIS'08)*, pp. 168–173, IEEE, Barcelona, Spain, September 2008.

[19] I. Leo, M. de Gemmis, P. Lops, G. Semeraro, and P. Molino, "Can a recommender system induce serendipitous encounters?" in *E-commerce*, pp. 239–245, InTech, London, UK, 2010.

[20] P. Adamopoulos and T. Alexander, "On unexpectedness in recommender systems: or how to better expect the unexpected," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 5, no. 4, pp. 1–32, 2014.

[21] S. Makri and A. Blandford, "What is serendipity? A workshop report," *Information Research*, vol. 16, no. 3, p. 491, 2011.

[22] A. Pease, S. Colton, R. Ramezani, J. W. Charnley, and K. Reed, "A discussion on serendipity in creative systems," in *Proceedings of the Fourth International Conference on Computational Creativity*, pp. 64–71, The University of Sydney, Sydney, Australia, June 2013.

[23] L. McCay-Peet and E. G. Toms, "Investigating serendipity: how it unfolds and what may influence it," *Journal of the Association for Information Science and Technology*, vol. 66, no. 7, pp. 1463–1476, 2015.

[24] S. Frenkel, M. Korczynski, L. Donoghue, and K. Shire, "Reconstituting work: trends towards knowledge work and infonormative control," *Work, Employment & Society*, vol. 9, no. 4, pp. 773–796, 1995.

[25] T. H. Davenport, *Thinking for a Living: How to Get Better Performances and Results from Knowledge Workers*, Harvard Business Press, Boston, MA, US, 2005.

[26] D. Skyrme, *Knowledge Networking: Creating the Collaborative Enterprise*, Routledge, London, UK, 2007.

[27] K. Valkokari, "Business, innovation, and knowledge ecosystems: how they differ and how to survive and thrive within them," *Technology Innovation Management Review*, vol. 5, no. 8, pp. 17–24, 2015.

[28] M. G. Russell, J. Huhtamäki, K. Still, N. Rubens, and R. C. Basole, "Relational capital for shared vision in innovation ecosystems," *Triple Helix: A Journal of University-Industry-Government Innovation and Entrepreneurship*, vol. 2, no. 1, p. 36, 2015.

[29] G. Schreyögg and J. Sydow, "Crossroads—organizing for fluidity? Dilemmas of new organizational forms," *Organization Science*, vol. 21, no. 6, pp. 1251–1262, 2010.

[30] M. Arena, R. Cross, J. Sims, and M. Uhl-Bien, "How to catalyze innovation in your organization," *MIT Sloan Management Review*, vol. 58, no. 4, pp. 38–48, 2017.

[31] M. F. Bosenman, "Serendipity and scientific discovery," *The Journal of Creative Behavior*, vol. 22, no. 2, pp. 132–138, 1988.

[32] P. V. Andel, "Anatomy of the unsought finding: serendipity: orgin, history, domains, traditions, appearances, patterns and programmability," *The British Journal for the Philosophy of Science*, vol. 45, no. 2, pp. 631–648, 1994.

[33] R.-H. Liang, "Designing for unexpected encounters with digital products: case studies of serendipity as felt experience," *International Journal of Design*, vol. 6, no. 1, pp. 41–58, 2012.

[34] X. Sun, S. Sharples, and S. Makri, "A user-centred mobile diary study approach to understanding serendipity in information research," *Information Research*, vol. 16, no. 3, p. 492, 2011.

[35] X. Zhou, X. Sun, Q. Wang, and S. Sharples, "A context-based study of serendipity in information research among Chinese

scholars," *Journal of Documentation*, vol. 74, no. 3, pp. 526–555, 2018.

[36] J. Dewey, *Experience and Nature*, Literary Licensing, Whitefish, MO, USA, 2013.

[37] K. Martin and A. Quan-Haase, ""A process of controlled serendipity": an exploratory study of historians' and digital historians' experiences of serendipity in digital environments," *Proceedings of the Association for Information Science and Technology*, vol. 54, no. 1, pp. 289–297, 2017.

[38] R. M. C. de Melo, *Call to Adventure: Designing for Online Serendipity*, Master's thesis, Faculdade de Belas Artes, Universidade do Porto, Porto, Portugal, 2012.

[39] C. Lutz, C. Pieter Hoffmann, and M. Meckel, "Online serendipity: a contextual differentiation of antecedents and outcomes," *Journal of the Association for Information Science and Technology*, vol. 68, no. 7, pp. 1698–1710, 2017.

[40] L. Björneborn, "Three key affordances for serendipity: toward a framework connecting environmental and personal factors in serendipitous encounters," *Journal of Documentation*, vol. 73, no. 5, pp. 1053–1081, 2017.

[41] L. McCay-Peet and A. Quan-Haase, "The influence of features and demographics on the perception of twitter as a serendipitous environment," in *Proceedings of the 27th ACM Conference on Hypertext and Social Media*, pp. 333–335, ACM, Nova Scotia, Canada, July 2016.

[42] C. Lutz, M. Meckel, and G. Ranzini, "Trusted surprises?: antecedents of serendipitous encounters online," in *Proceedings of the 63rd Annual Conference of the ICA International Communication Association*, p. 32, ICA International Communication Association, London, UK, March 2013.

[43] C. Brown, C. Efstratiou, I. Leontiadis, D. Quercia, and C. Mascolo, "Tracking serendipitous interactions: how individual cultures shape the office," in *Proceedings of the 17th ACM conference on Computer Supported Cooperative Work & Social Computing*, pp. 1072–1081, ACM, Baltimore, MD, USA, February 2014.

[44] T. Erickson and W. A. Kellogg, "Social translucence: an approach to designing systems that support social processes," *ACM Transactions on Computer-Human Interaction*, vol. 7, no. 1, pp. 59–83, 2000.

[45] J. M. Mayer, "Is there a place for serendipitous introductions?" in *Proceedings of the Companion Publication of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pp. 73–76, ACM, Baltimore, MD, USA, February 2014.

[46] J. M. Mayer, Q. Jones, and S. R. Hiltz, "Identifying opportunities for valuable encounters: toward context-aware social matching systems," *ACM Transactions on Information Systems*, vol. 34, no. 1, pp. 1–32, 2015.

[47] N. Eagle and A. Pentland, "Social serendipity: mobilizing social software," *IEEE Pervasive Computing*, vol. 4, no. 2, pp. 28–34, 2005.

[48] J. C. Losada, W. Creixell, T. Arredondo, P. Olivares, and R. María Benito, "Serendipity in social networks," *Networks and Heterogeneous Media*, vol. 7, no. 3, pp. 363–371, 2012.

[49] Xi Niu and F. Abbas, "A framework for computational serendipity," in *Proceedings of the Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization*, pp. 360–363, ACM, Bratislava, Slovakia, July 2017.

[50] R. V. D. Hoff, *Mastering the Global Transition on Our Way to Society 3.0*, Lindonk & de Bres, Amsterdam, Netherlands, 2014.

[51] T. Olsson, J. Huhtamäki, and H. Kärkkäinen, "Directions for professional social matching systems," *Communications of the ACM*, vol. 63, no. 2, pp. 60–69, 2020.

[52] J. Saldaña, *The Coding Manual for Qualitative Researchers*, SAGE, Newcastle upon Tyne, UK, 2015.

[53] J. Bardzell and S. Bardzell, "Humanistic HCI," *Synthesis Lectures on Human-Centered Informatics*, vol. 8, no. 4, pp. 1–185, 2015.

[54] L. Terveen and D. W. McDonald, "Social matching: a framework and research agenda," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 12, no. 3, pp. 401–434, 2005.

[55] C.-H. Tsai and B. Peter, "Beyond the ranked list: user-driven exploration and diversification of social recommendation," in *Proceedings of the 23rd International Conference on Intelligent User Interfaces*, pp. 239–250, ACM, Tokyo, Japan, March 2018.

[56] M. McPherson, L. Smith-Lovin, and J. M. Cook, "Birds of a feather: homophily in social networks," *Annual Review of Sociology*, vol. 27, no. 1, pp. 415–444, 2001.

[57] D. Easley and J. Kleinberg, "Networks, crowds, and markets: reasoning about a highly connected world," *Significance*, vol. 9, pp. 43-44, 2012.

[58] M. S. Granovetter, "The strength of weak ties," *American Journal of Sociology*, vol. 6, no. 78, pp. 1360–1380, 1973.

[59] A. Pentland, "Beyond the echo chamber," *Harvard Business Review*, vol. 11, no. 91, p. 80, 2013.

[60] M. Del Vicario, G. Vivaldo, A. Bessi et al., "Echo chambers: emotional contagion and group polarization on facebook," *Scientific Reports*, vol. 6, no. 1, pp. 1–12, 2016.

[61] T. Leong, S. Howard, and F. Vetere, "Feature take a chance on me: using randomness for the design of digital devices," *Interactions*, vol. 15, no. 3, pp. 16–19, 2008.

[62] L. Tuck, *Understanding Serendipitous Experiences When Interacting with Personal Digital Content*, University of Melbourne, Department of Information Systems, Melbourne, Australia, 2009.

[63] L. Tuck, R. Harper, and T. Regan, "Nudging towards serendipity: a case with personal digital photos," in *Proceedings of the 25th BCS Conference on Human-Computer Interaction*, pp. 385–394, British Computer Society, Newcastle, UK, July 2011.

[64] J. Bardzell, "Critical and cultural approaches to HCI," in *The SAGE Handbook of Digital Technology Research*, pp. 130–143, SAGE, Newcastle upon Tyne, UK, 2013.

[65] K. K. Myers and K. Sadaghiani, "Millennials in the workplace: a communication perspective on millennials' organizational relationships and performance," *Journal of Business and Psychology*, vol. 25, no. 2, pp. 225–238, 2010.

[66] E. Olshannikova, T. Olsson, J. Huhtamäki, and H. Kärkkäinen, "Conceptualizing big social data," *Journal of Big Data*, vol. 4, no. 1, p. 3, 2017.

[67] J. M. Mayer, S. R. Hiltz, L. Barkhuus, K. Väänänen, and Q. Jones, "Supporting opportunities for context-aware social matching: an experience sampling study," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 2430–2441, ACM, San Jose, CA, USA, May 2016.

[68] D. Obstfeld, "Social networks, the tertius iungens orientation, and involvement in innovation," *Administrative Science Quarterly*, vol. 50, no. 1, pp. 100–130, 2005.

[69] A. Endert, M. S. Hossain, N. Ramakrishnan, C. North, P. Fiaux, and C. Andrews, "The human is the loop: new directions for visual analytics," *Journal of Intelligent Information Systems*, vol. 43, no. 3, pp. 411–435, 2014.

[70] N.-N. Zheng, Z.-Y. Liu, P.-J. Ren et al., "Hybrid-augmented intelligence: collaboration and cognition," *Frontiers of*

Information Technology & Electronic Engineering, vol. 18, no. 2, pp. 153–179, 2017.
[71] J. M. Corrigan, "Augmented intelligence—the new AI—unleashing human capabilities in knowledge work," in *Proceedings of the 34th International Conference on Software Engineering*, pp. 1285–1288, IEEE Press, Zurich, Switzerland, June 2012.
[72] B. Scassellati and K. M. Tsui, "Co-robots: humans and robots operating as partners: the confluence of engineering-based robotics and human-centered application domains," *Handbook of Science and Technology Convergence*, Springer, Cham, Switzerland, pp. 1–10, 2014.
[73] B. Schwartz, *The Paradox of Choice: Why More Is Less*, HarperCollins, New York, USA, Revised edition, 2009.
[74] L. McCay-Peet, E. G. Toms, and E. K. Kelloway, "Examination of relationships among serendipity, the environment, and individual differences," *Information Processing & Management*, vol. 51, no. 4, pp. 391–412, 2015.
[75] H. M. Lefcourt, *Locus of Control: Current Trends in Theory & Research*, Psychology Press, Hove, UK, 2014.

PUBLICATION
II

What supports serendipity on twitter? online survey on the role of technology characteristics and their use

Olshannikova, E., Pirkkalainen, H., Olsson, T., and Huhtamäki, J.,

25th International Academic Mindtrek conference (Academic Mindtrek 2022), November 16–18, 2022, Tampere, Finland, 13
https://doi.org/10.1145/3569219.3569346

What Supports Serendipity on Twitter? Online Survey on the Role of Technology Characteristics and Their Use

Ekaterina Olshannikova
ekaterina.olshannikova@tuni.fi
Tampere University
Tampere, Finland

Henri Pirkkalainen
henri.pirkkalainen@tuni.fi
Tampere University
Tampere, Finland

Thomas Olsson
thomas.olsson@tuni.fi
Tampere University
Tampere, Finland

Jukka Huhtamäki
jukka.huhtamaki@tuni.fi
Tampere University
Tampere, Finland

ABSTRACT

Serendipity experiences are highly desirable in work life, considering both individuals' learning and organizational innovation capacity. This study looks into information and social serendipity in the context of Twitter. While Twitter can be viewed as a fruitful platform for serendipity to emerge, there is little understanding of what technology characteristics and use practices contribute to such experiences in work-related use. Drawing from the functional affordances theory, the paper investigates the role of presenteeism, self-disclosure, recommendation quality and pace of change, and different types of Twitter use as possible antecedents of serendipity. A cross-sectional international online survey was conducted with 473 respondents who actively use Twitter in their work. An exploratory factor analysis was performed, followed by linear regression analysis to identify relevant statistical associations. The findings indicate that presenteeism (i.e., the fundamental element of reachability) seems to have an effect on serendipity while the more designable characteristics, like the quality of recommendations, do not. Overall, the findings imply that serendipity experiences are primarily explained by individual characteristics like personality and specific ways of using Twitter. This is amongst the first studies on the role of Twitter characteristics as functional affordances in the formation of serendipity. The extensive empirical study contributes a detailed analysis of the antecedents of serendipity and opens avenues for research and design to identify new serendipity-inducing mechanisms.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**; • **Information systems → Social networks**; *Internet communications tools*; *Web searching and information discovery*.

Academic Mindtrek 2022, November 16–18, 2022, Tampere, Finland

ACM ISBN 978-1-4503-9955-5/22/11.
https://doi.org/10.1145/3569219.3569346

KEYWORDS

Information serendipity, Social serendipity, Serendipitous Social Encounters, IT-supported Serendipity, Social networking, User experience, Knowledge work, Online survey

ACM Reference Format:
Ekaterina Olshannikova, Henri Pirkkalainen, Thomas Olsson, and Jukka Huhtamäki. 2022. What Supports Serendipity on Twitter? Online Survey on the Role of Technology Characteristics and Their Use . In *25th International Academic Mindtrek conference (Academic Mindtrek 2022), November 16–18, 2022, Tampere, Finland.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3569219.3569346

1 INTRODUCTION

Social media services are actively utilized in work life as they open access to vast networks of relevant knowledge and experts and enable the exchange of ideas and expertise [71]. Organizational studies demonstrated that employees strategically use micro-blogging services like Twitter to increase their visibility and influence through the expression and promotion of professional identities [73]. Prior research suggests that Twitter supports the discovery of unexpected yet valuable content and contacts for professional interests [56], which implies that Twitter use offers a relevant empirical context to study experiences of serendipity.

Serendipity is conceptualized as an unsought yet fortunate experience prompted by an individual's interaction with ideas, information, objects, or phenomena [47]. Prior research on serendipity seems to largely focus on its role in information retrieval [2, 6], its importance in creativity and innovation [3, 32], and in knowledge building and learning [9, 55]. In recent literature, serendipity has been recognized as a strongly positive experience worth pursuing [16, 37], especially in professional activities [52, 57]. Since this study focuses on serendipity within the context of Twitter use, we cover both *information* and *social* serendipity.

However, little is known about if and how the technology characteristics and especially designable service features might support serendipity experiences or if certain usage practices tend to lead to such. Despite its desirability as a specific user experience, designing for serendipity will remain an elusive goal if the contributing mechanisms are unclear. Since chance and happenstance are central elements of serendipity, the possible factors contributing to it are likely diverse. Prior work features only a few attempts to reveal the antecedents of IT-supported serendipity, primarily in the context of information retrieval [39, 46], with little attention to its social counterpart. Therefore, by investigating a range of factors, the present study extends the understanding of the IT characteristics that serve as antecedents of information serendipity or social serendipity on the most popular microblogging service Twitter. We considered Twitter a fruitful context for researching serendipity as it is commonly and strategically used for professional purposes, and serendipity is an important element in its user experience [59, 73].

Thus, this exploratory research focuses on the following questions: **RQ1.** *What technological characteristics contribute to the experiences of information and social serendipity on Twitter?* **RQ2.** *How do these characteristics associate with the types of Twitter use?* We conducted a cross-sectional online survey of 473 respondents who use Twitter in their work. Drawing from the functional affordances theory [25, 48], we analyzed the effects of perceived *recommendation quality* [54], *pace of change* [5], *self-disclosure* [76] and *presenteeism* [5], and their relations to experiences of serendipity. Additionally, we included background and personality characteristics as control variables to understand the overall proportion of service features in explaining serendipity.

The quantitative analysis demonstrates that the examined technology characteristics provide limited direct support for serendipity. Nonetheless, all the investigated characteristics appear to be significantly associated with different types of Twitter use. The findings also illustrate that users who consume content by following active discussions and exploring others' tweets experienced higher levels of information serendipity than users who were actively producing content. Additionally, personality characteristics, such as *openness to experience*, *neuroticism*, and *conscientiousness*, were found to be essential in serendipity.

The primary contribution of the article is the report of an extensive quantitative study on how various technology characteristics and types of Twitter use can support the emergence of information and social serendipity in the context of professional life. The results provide insights into technology's role in shaping serendipity. In contrast to prior research, which focuses on serendipitous encounters related to information discovery, we extend the understanding of the little-studied concept of social serendipity. Furthermore, we anticipate the study to encourage the exploration of new serendipity-inducing mechanisms on social media services utilized for professional purposes.

2 RELATED WORK

Many research fields studied serendipity, resulting in numerous conceptualizations 1. For instance, in organization studies, the concept is treated as a behavioral and social pattern worth pursuing due to the positive effects on knowledge work and collaboration [30, 59]. In computer science, the concept is relevant, especially in information retrieval (IR) and Recommender Systems (RecSys) research, where it is regarded as a measure for preventing algorithmic bias and enabling information diversity [36, 60].

Research in Human-Computer Interaction (HCI) and Computer-Supported Cooperative Work (CSCW) has explored two branches of the phenomenon: natural serendipity, which is unplanned and non-deterministic, and IT-supported serendipity—facilitated or triggered by technology. Studies of natural serendipity cover, for instance, exploratory studies on social awareness and impromptu encounters within work environments [34]. Research on IT-supported serendipity has focused on designing and evaluating IT artifacts that aim to facilitate chance encounters between co-workers or collocated individuals [58] and recommender systems for surprising content discoveries [1].

Search engines present a classic example of IT-supported serendipity by opening access to enormous content and different pathways to encounter it. Notably, Twitter and other services with user-generated content differ from traditional information search engines by exposing users to content that is not deliberately searched for. Additionally, since Twitter is used to support diverse tasks, it is found to make discoveries unpredictable [68]. Such multi-purpose use of the platform promotes one of the key qualities of serendipity—the revelation of unsought connections [20]. On Twitter, serendipity can emerge due to the dynamics of personal social networks and user-generated content, which are enabled via 'follow,' 'mention,' 'favorite,' and 'retweet' features [62].

For this study, it is relevant to establish a conceptual separation between two target branches of serendipity—information and social serendipity. Information serendipity refers to unexpected yet fortunate discoveries of information and manifests in stumbling upon useful content (e.g., tweets, links, and hashtags). Social serendipity refers to unexpected encounters with other people (e.g., followees and followers), resulting in personal or professional benefits. The following sections will further conceptualize these two facets of serendipity.

2.1 Information Serendipity

Prior research on information serendipity has discussed whether it is possible to design for serendipity at all. After all, technology features tailored to satisfy and predict users' desires might decrease the element of surprise [39, 42]. While engineering serendipity with technology is considered an oxymoron, researchers concluded that it is possible to enable experiences that can be subjectively perceived as serendipitous [4, 41]. For instance, Makri et al. [42] proposed design strategies such as facilitating the revelation of patterns, seizing opportunities, relaxing personal boundaries, and supporting making a connection with previous experiences.

The perception of serendipity depends on various factors, which can prevent or establish the context for serendipitous encounters, for instance, trust and privacy [39]. By studying accidental discoveries on Twitter, McCay-Peet and Quan-Haase [46] also revealed the key factors influencing the perception of serendipity like user's age and activity level. The older the user and the more active she is on the platform, the higher the probability of fortunate information discovery and the strength of the perception of serendipity.

Design-oriented research in this domain has produced various artifacts, mainly content recommender systems, that support surprising discoveries. For instance, Toms and McCay-Peet (2009) designed and evaluated "a serendipity inducing tool" that enables unexpected suggestions from Wikipedia readings. Campos and Figueiredo [12] implemented a web search system, 'Max,' that allows divergent exploration of potentially useful Internet resources. Such systems typically utilize a similarity-based recommendation approach that relies on a history of users' inputs. More unexpected information

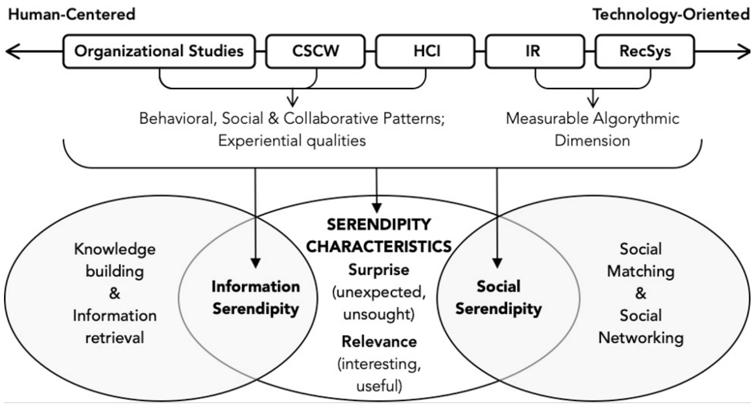


Figure 1: Overview of the serendipity conceptualization in different scientific fields, key characteristics of the serendipity, and the difference between information and social serendipity.

is provided in interactive library visualizations that support the opportunistic exploration of books [70]. Such services provide a more random and diverse pool of content selection, yet the chances of getting relevant suggestions are decreased.

2.2 Social Serendipity

While serendipity has been primarily studied within the context of information retrieval, little attention is paid to its social counterpart, that is, unexpected yet beneficial encounters with relevant people. In HCI and CSCW research, social serendipity is rarely addressed explicitly, even though related topics have been investigated. For instance, there are exploratory studies on daily spontaneous encounters [34], chance encounters [31], and social awareness [75] in collocated work environments [58]. A recent survey study highlights that not all chance encounters result in serendipitous experiences and suggests that technology might play a minor role in the process of social serendipity [57].

Design-oriented studies have investigated the means to enable chance encounters (or impromptu encounters), aiming to enhance awareness and interactions between collocated or distributed workers. For instance, Erickson and Kellogg [22] introduced the concept of 'socially translucent systems,' which increases the visibility of employees' activities in the context of large groups and organizations. Jeffrey and McGrath [31] designed a collaborative working environment for informal online interactions to help employees to make new connections and share knowledge. Eagle and Pentland [21] built a socially curious mobile service, which senses the contextual surroundings and encourages face-to-face interactions within a proximity range. Interestingly, such solutions are typically based on maximizing similarity in the social matching process, which might lead to anticipated rather than serendipitous encounters.

In summary, prior research agrees that technology artifacts could provide favorable conditions for serendipity. However, the question of how information systems could facilitate serendipity experiences remains unanswered. In contrast to the preceding research, we seek to understand the characteristics that facilitate both information and social serendipity in the professional use of Twitter. To understand the role of technology in the overall scheme of possible factors, we also investigate types of Twitter use, followership statistics, and personal background characteristics.

2.3 Serendipity and Professional Use of Twitter

Recent research on the work-related use of Twitter has established that serendipity experiences are critical for Twitter users and vital for organizational innovation practices [43, 59]. In a professional context, serendipity occurring in social media was found to foster the creation and use of new ideas [30, 73] and positively influence employees' performance [13, 59].

As Twitter allows vocationally motivated interactions between various individuals, communities, companies, and markets [50, 69], previous work has identified different professional purposes for using Twitter. For example, van Zoonen et al. [73] provided a holistic overview of the various types of work-related communication patterns on Twitter, categorizing work-, profession-, organization-related tweets, and employee-public interactions. Table 1 summarizes the three main categories for the professional use of Twitter addressed in prior research. In the present study, we focus on the professional use of Twitter from the individuals' perspective and investigate whether the platform supports their work activities and professional networking.

To our best knowledge, serendipity in the professional use of Twitter is acknowledged in prior research but not empirically studied. While it has been established that individuals can experience work-related serendipity on Twitter [7, 17, 67, 77], little is known about what, in particular, enables it. In what follows, based on existing theories and frameworks, we provide a conceptual overview of the inherent characteristics of Twitter that may enable serendipity.

Table 1: Summarized categories of professional purposes of Twitter use.

Category	Description
Professional Identity Management [17, 56, 69, 73]	Identifying and promoting self as representative of an organization, team/group, or profession
Knowledge Sharing [18, 30, 73]	Documentation of daily work activities, sharing of professional opinions; engagement in professional discourse (e.g., question and answer type of communication)
Professional Networking [30, 69, 73]	Expert finding to fill in the knowledge gap; building connections with like-minded professionals within and beyond organizational boundaries; sustaining ties with co-workers

3 TECHNOLOGY CHARACTERISTICS OF TWITTER AS POTENTIAL FACILITATORS OF SERENDIPITY

According to Martin and Quan-Haase [43], the dynamism and liveliness of user-generated content and the various recommendation features embedded in the user interface enable divergent exploration of interesting others and trending topics. Building on this, we theorize that the technology characteristics of Twitter can enable both information and social serendipity. In the following, we operationalize Twitter technology characteristics by drawing from the perspective of functional affordances [25, 48]. Functional affordances are the system's characteristics that "help or aid the user in doing something" (in this study, achieving serendipity experiences in a professional context). To define technology characteristics and establish study measures, we adapt the framework by Ayyagari et al. [5] (see Table 2). While the investigated experiential phenomenon by Ayyagari et al. [5] was different, the level of abstraction in IT features analysis seemed appropriate for our analysis.

Table 2: The framework of technology characteristics and their manifestation in the context of Twitter use.

Features	Characteristic	Characteristic's manifestation on Twitter
Intrusive	*Presenteeism* – reachability and accessibility	Tweet, re-tweet, mention, 'like', 'follow', and direct messaging
	Self-disclosure – making the self known to others	Twitter profile information, tweets and likes
Usability	*Recommendation quality* – usefulness and relevance	Tweet-timeline, Explore, You might like, Who to follow and Trends for you features
Dynamic	*Pace of change* – frequency of changes in IT environment	Dynamism of the Twitter feed, users' actions and interactions, and dynamics of social network structures

Presenteeism is defined by Ayyagari et al. [5] as the degree to which technology makes people reachable for communication. The concept was primarily investigated as a cause for technostress, and a task disruption factor [5, 63]. However, as Brooks [8] demonstrated, presenteeism can also have positive effects when social media is used for gaining personal benefits. Presenteeism can thus be measured by the extent to which users perceive that technology makes them and other users reachable and accessible. In the context of Twitter use, it could be seen as a core characteristic that motivates the various uses of the service, hence providing chances for unexpectedly reaching useful content or contacts.

Self-disclosure relates to the system's capability to enable and encourage user profile creation and users' willingness to expose their information. Twitter has limited agency regarding how each profile turns out, while the user community introduces some norms. Prior research demonstrated that the core functionality of social networking sites necessitates extensive self-disclosure [76]. The user's perspective refers to managing online connections, knowledge, and opinion sharing. From the service perspective, self-disclosure enables personalization—delivering relevant recommendations and pushing specific content. On Twitter, self-disclosure is mainly enabled via user profiles where actors may reveal relevant information about themselves. The characteristic can be measured by how users perceive their Twitter profile as descriptive, comprehensible, and up-to-date.

By suggesting actions to the user, **recommendation quality** refers to the system's most proactive (high-agency) features. It is widely studied in computer science, outlining dimensions that constitute useful recommendations [54]. The first dimension—*accuracy*—stands for the recommendation agent's capability to predict the user's preferences [28]. Next, the *novelty* dimension refers to producing surprise—recommendations beyond the typical users' interest [74]. Finally, *diversity*—delivers heterogeneous recommendations to overcome the filter bubble and avoid monotonous suggestions [53]. The characteristic could be measured via subjective perceptions regarding content relevance and contact recommendations.

Pace of change is defined by Ayyagari et al. [5] as subjective perceptions regarding the rapidness of changes within the service environment. This characteristic builds on the users' actions and interactions within the service, enabling dynamism and liveliness. On Twitter, this characteristic is manifested through the fast-paced changes in users' feed, their interactions, and changes in the structures of ego-centric social networks caused by followership activities. Thus, this characteristic can be measured by the extent that users perceive the frequency of changes on Twitter.

4 METHODOLOGY

To address the research questions with a quantitative approach, we ran an international online survey in English. The survey enabled the collection of a diverse sample of responses to an extensive number of Likert-scale questions.

4.1 Measures Used in the Study

All the survey items except *professional use* were adapted from existing and validated scales. Please see the full list of the survey items in Table 3. **(1) Serendipity**, including *information serendipity* (see IS1-IS3 items in Table 3 and *social serendipity* (SS1-SS3) items were adapted from Lutz et al. scale [40]. We adjusted the scale

Table 3: Constructs and indicators of the study, including serendipity, Twitter characteristics, types of Twitter use, and personality characteristics.

Constract	α	Mean	SD	Item	Loading
				SERENDIPITY	
Social Serendipity	0.81	5.25	1.11	**SS1.** When using Twitter, I have made an accidental fortunate discovery of a contact that was useful for me	**0.716*****
				SS2. When using Twitter, I have encountered useful contacts that I was not looking for	**0.635*****
				SS3. When using Twitter, I have made an unexpected fortunate discovery of a contact that was useful for me	**0.608*****
Information Serendipity	0.86	5.09	1.19	**IS1.** When using Twitter, I have made an accidental fortunate discovery of content that was useful for me	**0.655*****
				IS2. When using Twitter, I have made an unexpected fortunate discovery of content that was useful for me	**0.631*****
				IS3. When using Twitter, I have encountered useful information, ideas, or resources that I was not looking for	**0.430*****
				TWITTER CHARACTERISTICS	
Presenteeism	0.84	5.61	0.89	**PRE1.** Twitter enables me to access others	**0.793*****
				PRE2. Twitter makes me accessible to others	**0.754*****
				PRE3. The use of Twitter enables others to have access to me	**0.708*****
				PRE4. The use of Twitter enables me to be in touch with others	**0.639*****
Self-disclosure	0.81	4.80	1.17	**SD1.** My Twitter profile contains all data asked by the service	**0.731*****
				SD2. My Twitter profile says a lot about me	**0.701*****
				SD3. My Twitter profile is comprehensive	**0.679*****
				SD4. My Twitter profile is up-to-date	**0.611*****
Recommendation quality	0.88	4.75	1.15	**RQ1.** The recommended content on Twitter fits my preferences	**0.859*****
				RQ2. The recommended contacts to follow on Twitter are relevant to me	**0.831*****
				RQ3. The recommended content on Twitter is relevant to me	**0.762*****
				RQ4. The recommended contacts to follow on Twitter fit my preferences	**0.726*****
Pace of change	0.70	4.74	1.01	**PC1.** There are frequent changes in the feed of other users' tweets	**0.826*****
				PC2. The users whose tweets I see in my feed changes frequently	**0.516*****
				PC3. The topics in my Twitter feed change frequently	**0.507*****
				TYPES OF TWITTER USE	
Professional use	0.79	3.77	0.90	**PU1.** I use Twitter to support my work activities	**0.940*****
				PU2. I use Twitter in my work	**0.698*****
				PU3. I use Twitter to support professional networking	**0.523*****
Receiving	0.77	3.57	0.73	**REC1.** Follow discussions related to particular hashtags	**0.767*****
				REC2. Check trending hashtags	**0.624*****
				REC3. Look for new Twitter users to follow	**0.612*****
				REC4. Read other people's tweets	**0.492*****
Broadcasting	0.83	3.43	0.76	**BR1.** Add photos and videos to tweets	**0.886*****
				BR2. Add hashtags to tweets	**0.669*****
				BR3. Send your own tweets	**0.524*****
				BR4. Mention other Twitter users in tweets	**0.445*****
				BR5. Share public content from other digital media in Twitter	**0.428*****
Interacting	0.80	3.25	0.81	**INT1.** Discuss with other Twitter users via Tweets	**0.844*****
				INT2. Engage in dialogue with other Twitter users	**0.795*****
				INT3. Discuss with other Twitter users via direct messages	**0.601*****
				PERSONALITY CHARACTERISTICS	
Openness to experience	0.84	5.44	1.20	... imaginative	**0.869*****
				... creative	**0.822*****
Neuroticism	0.76	3.75	1.39	... anxious	**0.778*****
				... easily upset	**0.774*****
				... moody	**0.620*****
Agreeableness	0.77	5.49	0.96	... warm	**0.778*****
				... kind	**0.709*****
				... sympathetic	**0.654*****
Conscientiousness	0.70	5.34	1.13	... organized	**0.800*****
				... self-disciplined	**0.657*****
Extraversion	0.73	4.31	1.41	... extraverted	**0.783*****
				... talkative	**0.736*****

to the context of work-related Twitter use by reflecting both the information and social perspectives; **(2) Twitter characteristics**, including *presenteeism*, *self-disclosure*, *recommendation quality*, and *pace of change. Presenteeism* items (PRE1-PRE4) adapted from Ayyagari et al. [5]; *self-disclosure* (SD1-SD4 items) from Lutz et al. [40]. *Recommendation quality* items (RQ1-RQ4 in Table 3) were taken from Knijnenburg et al. [35] scale, which we adjusted by distinguishing between content and contact recommendations. *Pace of change* (items PC1-PC3) originate from the scale of the dynamism of content by Heide and Weiss [26]; **(3) Types of Twitter use**, including *professional use*, *receiving* (passive consumption of social media content), *broadcasting* (active creation of the social media content), *interacting* (active engagement between users of social media). *Professional use* variable (PU1-PU3 items) represents the extent to which Twitter was used to accomplish work in general.

These items were designed by the authors and also used for screening. For the remaining items, the scale of active vs. passive use of online social networking sites was adapted [10, 11, 44]; **(4) Personality characteristics**, adapted from The Big Five Personality Traits [24, 38, 66], including *openness to experience*, *neuroticism*, *agreeableness*, *conscientiousness*, *extroversion*.

In general, we adjusted the scales to the context of the work-related use of Twitter. All the items were measured on a Likert-scale ranging from 1 (strongly disagree) to 7 (strongly agree) except for the types of Twitter use scales, which measured the frequency of use ranging from 1 (Never) to 5 (Every time).

4.2 Recruitment and respondents

We implemented the survey in English and used the Prolific[1] service for recruiting respondents in May 2018. We targeted individuals who use Twitter for professional purposes. To understand serendipitous experiences broadly, we did not limit to specific professions or industries. To ensure the validity of the sample we requested Prolific to screen their entire panel for eligible individuals who (1) worked in part- or full-time positions and (2) used Twitter for work (e.g., to support work tasks or professional networking) on a daily, weekly, or monthly basis. Only 1,080 of 122,435 from respondents pool fulfilled these two criteria. We started to invite people from that pre-screened pool to participate in the study. Within nine days we collected answers from 546 respondents.

Next, we took several extra measures to ensure the validity of the sample. First, we included survey questions on the professional use of Twitter to ensure that the respondents represent the intended population. Second, as we utilized Likert scale statements, including reversed statements, the responses with a standard deviation lower than 0.5 across all the questions were removed as a probable indication of inattentive responding.

Further, invalid responses were detected in open text fields, and the respondents who provided unrelated or inappropriate comments were omitted from the data. We also excluded responses given in less than three minutes as, on average, filling in the survey took 9 minutes. These filtering actions resulted in 473 responses that we considered valid for the analysis. The characteristics of the studied sample are presented in Figure 2. Gender, age, Twitter use experience, and followership information was also used to analyze serendipity antecedents.

4.3 Data Analysis

We used IBM SPSS Statistics[2] version 24 for all the analyses. First, we conducted an exploratory factor analysis (principal axis factoring[3]) to construct the latent variables (see Table 3). The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy confirmed that the factoring of the items was possible with a value of 0.930 and a statistically significant Bartlett's test of sphericity (0.000). The extracted factors were allowed to correlate with each other by using the Promax rotation. Most of the item loadings (17/21, Table 3) were greater than 0.700, the threshold suggested by Chin [14], whereas the lowest loading was 0.574. Factor loadings greater than 0.400 have been considered acceptable in prior methodology literature [23]. We only omitted the following three items from three separate Big Five Personality Trait scales due to factor loadings lower than 0.400: "I see myself as ...unconventional (openness to experience-scale), ...dependable (conscientiousness-scale), ...enthusiastic (extraversion-scale)." The descriptive statistics of the factor analysis are shown in Table 3. The internal consistency of the scales was sufficient, as indicated by Cronbach's alpha value, which is >0.70 [65].

Second, we constructed three linear regression models to study the antecedents of information and social serendipity. We chose to preserve the range of the original survey items and calculated sum variables based on the means of the variables instead of using factor loadings as regression weights. The first model included gender, the age of respondents, and the number of followers and followees on Twitter. The second model extended the first model with five personality characteristics (*openness to experience, neuroticism, agreeableness, conscientiousness, extroversion*). The third model included all variables of the study.

Third, to analyze the Twitter use types, we constructed three linear regression models that included background, personality, and Twitter characteristics as antecedents of different types of Twitter use—*receiving, broadcasting, and interacting*.

5 FINDINGS

We first report descriptive statistics on the constructs to provide an overview of the data. Next, we provide findings on the factors that can explain information and social serendipity. Finally, we describe which background, personality, and Twitter characteristics are associated with different types of Twitter use.

5.1 Descriptive Statistics

Most respondents reported that they had experienced unexpected yet fortunate discovery of both content and contacts (See Figure 3). A similar positive attitude is also visible regarding Twitter characteristics. Respondents agree that Twitter makes them and other users accessible (presenteeism) and report their Twitter profile as complete, comprehensive, and up-to-date (self-disclosure). They also perceive the recommendations of both content and contacts on Twitter as effective (recommendation quality), and that Twitter is a dynamic platform (pace of change). As for personality characteristics, a large proportion of the sample reported being open to experience, agreeable, and conscientious. Attitude regarding being neurotic and extrovert is more evenly distributed, resulting in a median score of 4 (neutral).

The descriptive statistics on the types of Twitter use imply that a large portion of the sample represents active Twitter users (See Figure 4). Most respondents reported that they often use Twitter for professional purposes, for instance, to support work activities and professional networking (Med=4). The respondents reported almost equal frequencies of consuming (*receiving*, Med=4) and producing (*broadcasting*, Med=4) content on Twitter. At the same time, the respondents seem to less frequently engage in dialogue and discussions with other Twitter users (*interacting*, Med=3).

[1]Service for the online participant recruitment – https://www.prolific.co
[2]SPSS is a widely used program for statistical analysis – https://www.ibm.com/products/spss-statistics
[3]Principal axis factoring (PAF) that seeks the least number of factors that can account for the common variance (correlation) of a set of variables

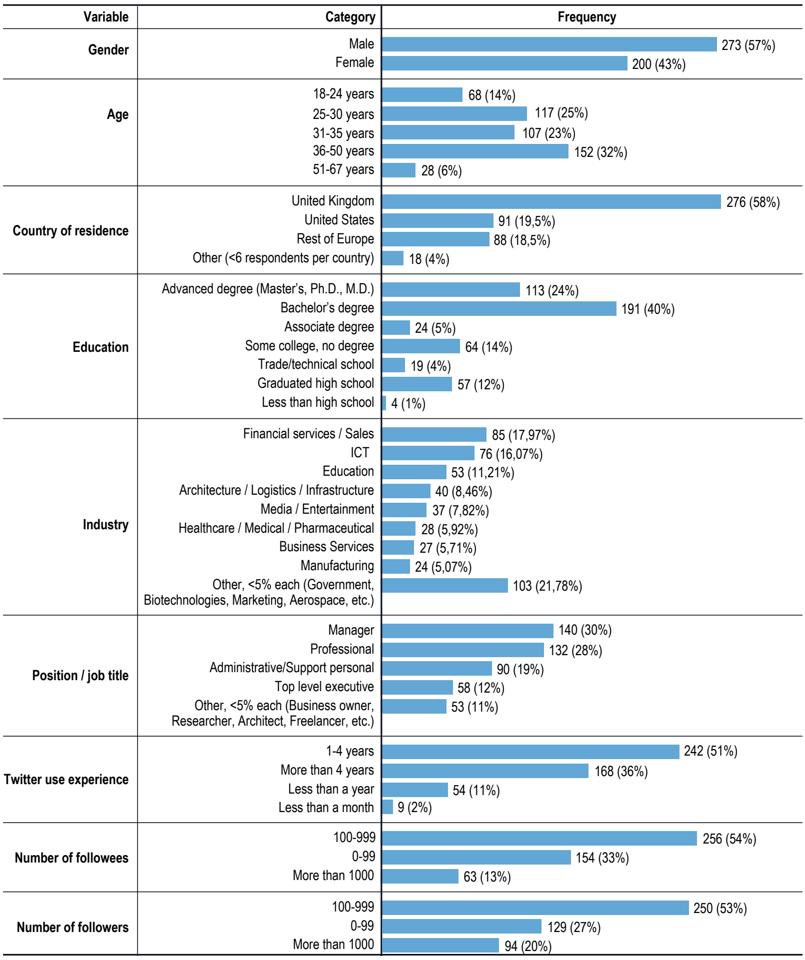


Figure 2: Sample characteristics (N = 473)

5.2 Antecedents of Information and Social Serendipity

Three linear regression models were utilized to reveal the antecedents of information and social serendipity on Twitter (See Table 4). From the first model (IS I & SS I), it is evident that the experience of Twitter use (0.223***), along with the number of followees (0.156*), associates with both information and social serendipity. It seems that the more the user is exposed to content on Twitter (over time and through followees), the higher are the chances for serendipity. The amount of variance extracted by the first model was low both in the case of information (3%) and social (2%) serendipity. That said, this effect does not explain the phenomenon well, which calls for investigating other factors.

The second model (IS II & SS II) adds two personality characteristics that can potentially associate with serendipity. This model performed slightly better regarding the amount of variance extracted as the model explained 13% of information serendipity and 11% of social serendipity. Accordingly, the more imaginative and creative (openness to experience, 0.272***), anxious, and moody (neuroticism, 0.070**) the user is, the higher the probability of encountering unexpected information. On the other hand, neuroticism does not have a significant effect (0.067) on social serendipity, while

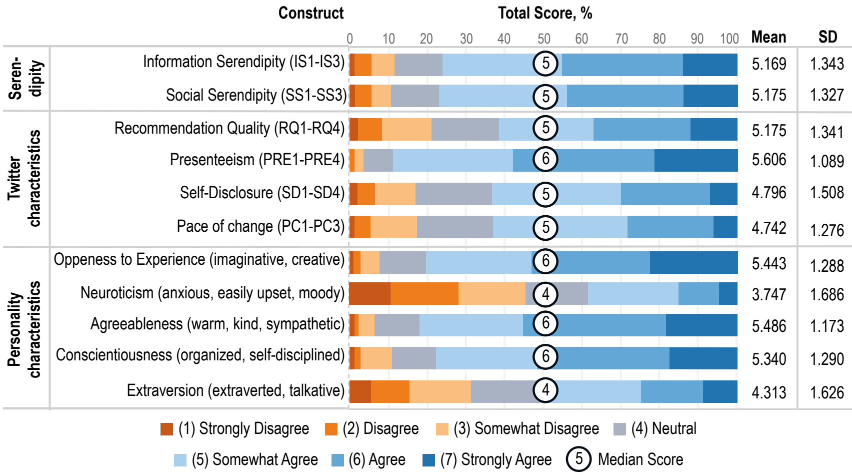


Figure 3: Descriptive statistics of the score distribution per constructs of serendipity, Twitter, and personality characteristics (%). N=473.

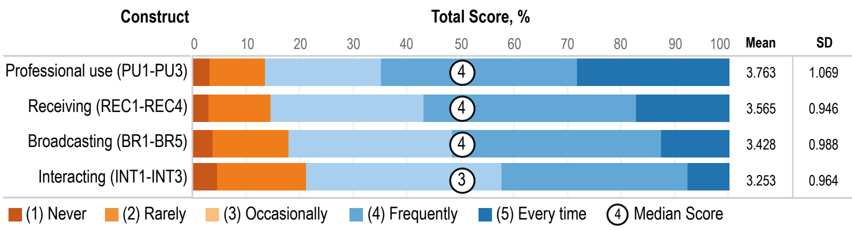


Figure 4: Descriptive statistics of the score distribution per constructs of Twitter use types (%). N=473.

agreeableness does (0.120*). The more warm, kind, and sympathetic the person is, the higher the probability of social serendipity seems.

The third model (IS III & SS III) explained 37% of information serendipity and 27% of social serendipity overall variance. When more variables are accounted for, the number of followees is no longer a significant predictor (0.89), and the number of followers is negatively associated with both information (-0.130*) and social serendipity (-0.161*). This suggests that individuals with a higher number of followers were less likely to perceive information and social serendipity than those with fewer followers. This might further consolidate the idea that mere exposure does not explain serendipity. The conscientiousness personality characteristic is similarly negatively associated with both information (-0.074*) and social serendipity (-0.113**). The less organized and self-disciplined the user, the higher the probability of encountering information and contacts serendipitously.

As for Twitter characteristics, only the extent to which users consider Twitter is making them reachable and accessible (presenteeism) seems to have a positive effect on information (0.593***) and social serendipity (0.508***), while recommendation quality, self-disclosure, and pace of change had no effect. This implies that many of the current Twitter features fail to support the emergence of this type of user experience—at least directly.

Using Twitter for professional purposes had a positive effect on both information (0.181***) and social serendipity (0.144**). The extent of using Twitter for passive actions (receiving (0.191**), e.g., reading tweets, checking hashtags) also has a positive effect on information serendipity. Engaging in discussions with other Twitter users (interacting) was negatively associated with information serendipity (-0.252***). This suggests that, within our sample, conversations with other Twitter users reduced the likelihood of perceiving unexpected fortunate discoveries of content.

Table 4: Antecedents of Information Serendipity (IS) and Social Serendipity (SS). N=473, unstandardized β values, CI= confidence intervals, * p <0.05, ** p <0.01, *** p <0.001

	IS I	SS I	IS II	SS II	IS III	SS III	95% CI	
	β (sig.)	β (sig.)	β (sig.)	β (sig.)	β (sig.)	β (sig.)	IS III	SS III
(Constant)	**4.365*****	**4.344****	**2.194*****	**2.147*****	.326	0.569	[-0.544, 1.195]	[-0.439, 1.577]
BACKGROUND CHARACTERISTICS								
Gender	-0.025	0.084	0.052	0.166	0.037	0.104	[-0.131, 0.205]	[-0.091, 0.298]
Age	-0.002	-0.006	-0.003	-0.006	-0.005	-0.008	[-0.014, 0.004]	[-0.018, 0.002]
Twitter use experience	**0.223*****	**0.154***	**0.201*****	**0.130***	**0.167*****	0.103	[0.047, 0.287]	[-0.036, 0.242]
Number of Followees	**0.156***	**0.195***	**0.171***	**0.199****	0.89	0.128	[-0.276, 0.016]	[-0.044, 0.299]
Number of Followers	0.004	0.028	-0.99	-0.080	**-0.130***	**-0.161***	[-0.059, 0.237]	[-0.330, 0.008]
PERSONALITY CHARACTERISTICS								
Openness to experience			**0.272*****	**0.253*****	**0.147*****	**0.158*****	[0.078, 0.234]	[0.067, 0.249]
Neuroticism			**0.070****	0.067	**0.080****	**0.075****	[0.015, 0.141]	[0.002, 0.148]
Agreeableness			0.087	**0.120***	-0.035	-0.018	[-0.144, 0.055]	[-0.134, 0.097]
Conscientiousness			0.029	-0.007	**-0.074***	**-0.113****	[-0.151, 0.012]	[-0.207, -0.018]
Extroversion			-0.002	0.041	0.001	0.025	[-0.065, 0.062]	[-0.049, 0.098]
TWITTER CHARACTERISTICS								
Presenteeism					**0.593*****	**0.508*****	[0.472, 0.696]	[0.378, 0.638]
Self-disclosure					-0.017	0.046	[-0.115, 0.080]	[-0.067, 0.159]
Recommendation quality					0.067	0.064	[-0.018, 0.162]	[-0.040, 0.169]
Pace of change					-0.031	-0.004	[-0.125, 0.077]	[-0.121, 0.113]
TYPES OF TWITTER USE								
Professional use					**0.181*****	**0.144****	[0.072, 0.286]	[0.020, 0.268]
Receiving					**0.191****	-0.038	[0.000, 0.318]	[-0.222, 0.146]
Broadcasting					0.060	0.007	[-0.084, 0.235]	[-0.179, 0.192]
Interacting					**-0.252*****	0.044	[-0.385, -0.108]	[-0.116, 0.205]
R squared	0.03	0.02	0.13	0.11	0.37	0.27		

5.3 Antecedents for Different Types of Twitter Use

As the different types of Twitter use have an effect on information serendipity especially, we investigated how the types of use might be associated with the other constructs (See Table 5).

The regression analysis results illustrate that the first and second models explain 40% of the overall variance in receiving and broadcasting, whereas the third model explains 34% of interacting. The data showcases a few background characteristics that might affect specific types of Twitter use. For instance, the more followees a user has, the more open she is to receiving (0.150***) behavior such as reading tweets, following and exploring hashtags, and seeking new Twitter users to follow. There is also a positive effect of the number of followers on broadcasting (0.225***) and interacting (0.166***). It is noteworthy that the findings only indicate a correlation, not a causal relation. One possible explanation for this correlation could be that the more followers one has, the more eager the user is to create and share content, mention others, and engage in discussions. Alternatively, the more one has shared content and interacted over time, the more followers they have managed to accumulate.

Openness to experience is associated with the receiving type of Twitter use (0.074***): the more open the person is, the higher the probability for her to be interested in exploring new information and contacts. Findings also illustrate that the less agreeable a person is, the more open she is to types of use such as receiving (-0.102***) and interacting (-0.071*). *Conscientiousness* had a relatively weak but positive association with interacting (0.053*). The more organized and self-disciplined a person is, the more open she is to discuss with other Twitter users. Furthermore, the personality characteristic of extroversion has a positive effect on receiving (0.044**), broadcasting (0.081***), and interacting (0.079**).

As for the Twitter characteristics, they all have an effect on different types of Twitter use. *Presenteeism* has a slightly lower effect on types of use than the other Twitter characteristics, as demonstrated by the positive effects on receiving (0.070*), *broadcasting* (0.064*), and *interacting* (0.078*). *Self-disclosure* has a positive effect on *receiving* (0.085***), *broadcasting* (0.196***), and *interacting* (0.183***), suggesting that exposing information about oneself is critical to all three forms of using Twitter. *Recommendation quality* seems to positively associate with types of Twitter use in terms of frequency of *receiving* (0.104***) and *interacting* (0.117***), yet it does not associate with serendipity directly. Furthermore, the *pace of change* (e.g., of Twitter feed) has an effect on *receiving* (0.246***), *broadcasting* (0.113***), and *interacting* (0.079**).

In summary, the findings imply that while Twitter characteristics were found not to have a direct statistical effect on serendipity, they associated with the ways of using the service, which in turn seem to support the emergence of serendipity. In other words, Twitter characteristics seem to relate to serendipity indirectly.

Table 5: Background, Personality, and Twitter Characteristics as antecedents of the Twitter use types. N=473, unstandardized β values, CI= confidence intervals * p <0.05, ** p <0.01, *** p <0.001

	Receiving β (sig.)	95% CI	Broadcasting β (sig.)	95% CI	Interacting (sig.)	95% CI
(Constant)	**0.636****	[0.078, 1.194]	**0.504***	[-0.073, 1.081]	0.213	[-0.341, 0.955]
BACKGROUND CHARACTERISTICS						
Gender	-0.065	[-0.172, 0.043]	**-0.114****	[-0.225, -0.003]	0.039	[-0.086, 0.164]
Age	-0.001	[-0.006, 0.005]	-0.003	[-0.008, 0.003]	-0.001	[-0.006, 0.007]
Twitter use experience	0.025	[-0.052, 0.103]	0.031	[-0.050, 0.111]	0.006	[-0.084, 0.096]
Number of Followees	**0.150*****	[0.056, 0.244]	0.000	[-0.097, 0.098]	0.070	[-0.039, 0.180]
Number of Followers	-0.046	[-0.136, 0.045]	**0.225*****	[0.131, 0.318]	**0.166*****	[0.061, 0.272]
PERSONALITY CHARACTERISTICS						
Openness to experience	**0.074*****	[0.024, 0.124]	0.039	[-0.013, 0.090]	0.001	[-0.057, 0.059]
Neuroticism	0.014	[-0.026, 0.055]	0.013	[-0.029, 0.055]	0.001	[-0.046, 0.048]
Agreeableness	**-0.102*****	[-0.165, -0.038]	-0.053	[-0.119, 0.012]	**-0.071***	[-0.145, 0.003]
Conscientiousness	0.034	[-0.018, 0.087]	0.034	[-0.021, 0.088]	**0.053***	[-0.008, 0.114]
Extroversion	**0.044****	[0.003, 0.084]	**0.081*****	[0.039, 0.122]	**0.071*****	[-0.024, 0.118]
TWITTER CHARACTERISTICS						
Presenteeism	**0.070***	[-0.002, 0.141]	**0.064***	[-0.010, 0.137]	**0.078***	[-0.005, 0.161]
Self-disclosure	**0.085*****	[0.024, 0.145]	**0.196*****	[0.134, 0.259]	**0.183*****	[0.113, 0.252]
Recommendation quality	**0.104*****	[0.047, 0.161]	0.046	[-0.013, 0.105]	**0.117*****	[0.050, 0.183]
Pace of change	**0.246*****	[0.185, 0.307]	**0.113*****	[0.050, 0.176]	**0.079****	[0.009, 0.150]
R squared	0.40		0.40		0.34	

6 DISCUSSION

This study investigated technology characteristics that can contribute to serendipity experiences and how they associate with Twitter use types. The findings provide several empirical contributions relevant to research on serendipity, computer-supported cooperative work, and information systems:

(1) We consolidate the relevance of the serendipity phenomenon within the context of Twitter use for professional purposes – respondents reported that they experienced both information and social serendipity. While the types of Twitter use in relation to serendipity were acknowledged and tentatively studied in prior research [7, 46, 77], our findings extend the empirical understanding of existing statistical associations, particularly in the context of professional use of Twitter. The use of Twitter for professional purposes was strongly associated with serendipity: it is about one's orientation toward or the practice of using it for specific purposes.

(2) This is the first study that investigates Twitter characteristics as functional affordances of serendipity, demonstrating technology's limited role in shaping serendipity. According to this sample and analysis, serendipity is more explained by factors relating to individual characteristics, such as personality and behavior, that is, how Twitter is used in work.

(3) In contrast to prior research that focuses solely on information serendipity [39, 46], we extend the understanding of social serendipity. The analysis contrasts information serendipity and social serendipity by identifying differences between factors that explain them. The findings generally show that the regression models explain more variance in information serendipity than social serendipity.

(4) In contrast to prior research, which only admits the importance of personality traits in serendipity [46], our study offers extensive empirical data on the associations between serendipity and personality characteristics in the context of Twitter use.

In what follows, we elaborate on the specific takeaway findings by reflecting on the study's contributions. We conclude by stating the limitations and pointing out practical implications and opportunities for future research.

6.1 Discussion of the Key Findings

6.1.1 The Role of Twitter Characteristics. To our knowledge, this is the first study that investigates Twitter characteristics as functional affordances that might contribute to serendipity experiences. We assume that the technology characteristics on Twitter (especially *recommendation quality*) are intended to enrich the user experience in many ways, including the facilitation of serendipity. However, our findings imply that the studied characteristics, apart from *presenteeism*, fail to do that, at least through a direct statistical association.

A possible explanation for such a finding could be the Twitter mechanisms of pushing the content to the users. Notably, the Twitter feed was previously organized in reverse-chronological order and only featured the followees' tweets. Nowadays, the feed displays the tweets in a personalized manner according to inferred user preferences [51, 72]. Even if a user does not have many followees, the system will provide an endless stream of content based on the recommender system (also from outside one's connections). Furthermore, the higher the number of followees users have, and the more active they are, the higher the probability that recommendations will be based on their established interest sphere. This, in turn, might result in "echo chambers" rather than serendipity, leading to interactions with similar ideologies and thoughts with fewer opportunities for new or controversial standpoints to emerge [19]. That said, the analysis indicates that Twitter characteristics are central antecedents to the different ways of using the platform. This implies that Twitter characteristics can be indirectly related to serendipity, at least regarding information serendipity.

6.1.2 The Role of Types of Twitter Use. Although it could be expected that using Twitter for *receiving* (e.g., following specific hashtags) and *interacting* (e.g., engaging in discussions) would support social serendipity, the data did not show such an effect. Different ways of using Twitter primarily associate with information serendipity. Especially the use of Twitter for *receiving* was found to increase the chances of experiencing information serendipity. Somewhat counter-intuitively, *interacting* with others by taking part in discussions seems to be negatively associated with information serendipity. We suggest that a possible reason for this negative relationship is that users in our sample engage in discussions with others who are already familiar with them. Hence, the user is less likely to encounter new information. This is because individuals with stronger ties, that is, they already are acquainted with each other and have a shared history of interactions, usually possess overlapping knowledge due to shared background and similar interests [19]. *Broadcasting* type of use does not associate with serendipity at all. Such findings imply that the more the platform knows about the user due to active content production, the more likely one ends up in a silo of like-minded actors—consolidating the theory of echo chambers [19, 53].

6.1.3 The Role of Personality Characteristics. The analysis shows that, regarding the personal factors, *openness to experience* and *neuroticism* are favorable for both information and social serendipity. The identified association between openness to experience and serendipity supports prior research that has conceptualized that people should have an open or "prepared" mind to experience serendipity [20, 45]. We suggest that *neuroticism* plays an essential role because neurotic people tend to mind-wander and shift their attention to task-irrelevant thoughts [64]. Such personality quality might be of benefit to experience serendipity because it can drive a person's ability to be receptive to unexpected connections and patterns, which is one of the suggested strategies to experience serendipity [42]. Here, we wish to note that *conscientiousness* (being self-disciplined and organized) may not be favorable for serendipity. Being conscientious means that a person would be task-oriented and have less space for idle moments and an associative state necessary for experiencing serendipity [42, 47]. Hughes et al. [29] suggest that the use of Twitter to socialize correlates with lower conscientiousness: the fact that average Twitter users possess a broad spectrum of interests and are keen on socializing would decrease the time for their goal-directed behavior.

6.2 Limitations

With the benefit of hindsight, we acknowledge some limitations in our study. First, the study was based on self-reported subjective experiences collected with a cross-sectional survey, which may not give an entirely truthful and generalizable picture of the actual serendipity experiences in the target context. Using Prolific to recruit respondents could bias the findings as such online platforms were criticized for providing poor-quality data and samples [15]. For instance, most of our respondents were from the United Kingdom and Western countries. There might be cultural factors related to the use of Twitter and the perception of serendipity that we wish to address in future research.

Despite the limitation of the sample, using Prolific enabled us to find suitable participants for the study, ensuring that each participant used Twitter for work and could answer the survey anonymously. Moreover, according to Newman et al. [49], Prolific provides a more diverse participant pool and ethical pricing for incentives compared to other online survey platforms. Additionally, we employed pre-screening and attention checks to ensure data quality, which are key recommendations for running research on platforms like Prolific [49].

Second, the study addressed serendipity in the context of professional use of Twitter with specific technological characteristics and user community. As such, we believe that the insights on the antecedents of serendipity may not be transferable to other social media services used in work- or non-work settings. Third, the participants were rather eager to consider their experiences serendipitous. With such a study setup, we could not control how strong serendipity experiences the participants had, and the measures could be said to set a relatively low threshold for what counts as serendipity.

6.3 Practical Implications and Future Research Topics

The results of this study can inform organizations and individuals who utilize Twitter for work purposes. Serendipity experiences have been found essential for task innovation in the context of Twitter use [61]. Thus, understanding serendipity antecedents can help organizations strategize the use of Twitter for professional purposes to induce serendipitous information discoveries and social networking. The findings also could inform new measurable algorithmic dimensions for researchers and designers in Information Retrieval and Recommender Systems fields. A minor role of technology in explaining serendipity calls for revisiting computational approaches and mechanisms on social media services. We encourage other researchers to pursue the identification of both internal characteristics and designable service features for inducing experiences of serendipity—particularly social serendipity. The findings reveal that the factors related to the individual (i.e., personality) play an essential role in using Twitter. These factors are harder to influence by design choices because they are inherent to the user. Instead, system design mindful of serendipity could aim to support such experiences with strategies that foster certain personality qualities responsive to the perception of serendipity experience [42]. The study also implies that influencing serendipitous experiences are hardly controllable by design choices. Therefore, rather than seeking design solutions that deliver "serendipity on a plate" [42], we call for design endeavors that aim to facilitate value-adding types of use. For instance, altering how information can be received or broadcasted or how conversations can be initiated with others might increase the chances of serendipity. That said, it is essential to set users in an explorative state of mind, and a promising approach in that regard is to apply diversity-enhancing strategies [27, 33] that can foster divergent exploration activities by pushing content and contacts that are beyond typical users' interests. Therefore, we encourage searching for new ways to boost interaction between individuals who are not well-acquainted with each other. This could potentially mitigate the unexpected finding that interacting with others reduces information serendipity. We also suggest that promoting different ways of using Twitter is helpful as it may indirectly lead to serendipity.

7 CONCLUSIONS

The emergence of serendipity can be regarded as a positive indicator of the overall service quality in many IT services and, therefore, a fundamental design goal. While Twitter can be considered a potent platform for serendipity, there is little understanding of what particular characteristics contribute to such experiences. After identifying potentially relevant Twitter characteristics, we collected and analyzed survey data to measure how the characteristics are associated with information and social serendipity experiences. The study offers an analysis of the factors contributing to serendipity in the context of the work-related use of Twitter. Additionally, we extend prior research by analyzing antecedents of serendipity to various technology and personality characteristics and types of Twitter use. The findings illustrate that serendipity is more explained by personality characteristics and types of Twitter use. Regarding Twitter characteristics, only presenteeism has a significant effect on subjective experiences of serendipity. Thus, this study illustrated the complex nature of serendipitous experiences in the professional use of Twitter, which calls for identifying new serendipity-inducing mechanisms.

ACKNOWLEDGMENTS

This work was conducted under Business Finland Project Big Match (3166/31/2017 and 3074/31/2017).

REFERENCES

[1] Panagiotis Adamopoulos and Alexander Tuzhilin. 2014. On Unexpectedness in Recommender Systems: Or How to Better Expect the Unexpected. *ACM Trans. Intell. Syst. Technol.* 5, 4, Article 54 (dec 2014), 32 pages. https://doi.org/10.1145/2559952

[2] Naresh Kumar Agarwal. 2015. Towards a definition of serendipity in information behaviour. *Information Research: An International Electronic Journal* 20, 3 (2015), 1–16.

[3] Theresa Dirndorfer Anderson. 2011. Beyond Eureka Moments: Supporting the Invisible Work of Creativity and Innovation. *Information Research: An International Electronic Journal* 16, 1 (2011), 1–24.

[4] Paul André and MC Schrafel. 2009. Computing and chance: designing for (un) serendipity. *The Biochemist E-Volution* 31, 6 (2009), 19–22.

[5] Ramakrishna Ayyagari, Varun Grover, and Russell Purvis. 2011. Technostress: technological antecedents and implications. *MIS quarterly* 35, 4 (2011), 831–858.

[6] David Bawden. 2011. Encountering on the road to Serendip? Browsing in new information environments. In *Innovations in Information Retrieval: Perspectives for Theory and Practice*, Alen Foster and Pauline Rafferty (Eds.). Facet Publishing, London, UK, 1–22.

[7] Toine Bogers and Lennart Björneborn. 2013. Micro-serendipity: Meaningful coincidences in everyday life shared on Twitter. In *iConference 2013 Proceedings, Fort Worth, Texas, USA, 12-15 February, 2013*. iSchools, Grandville, 196–208.

[8] Stoney L. Brooks. 2015. Being Social isn't Just About Fun: An Examination of Personal Social Media Usage. In *AMCIS 2015. Proceedings of the Twenty-first Americas Conference on Information Systems, Puerto Rico, USA, 13-15 August 2015*. Association for Information systems (AIS), Atlanta, Georgia, US.

[9] Ilona Buchem. 2011. Serendipitous learning: Recognizing and fostering the potential of microblogging. *Form@ re-Open Journal per la formazione in rete* 11, 74 (2011), 7–16.

[10] Moira Burke, Robert Kraut, and Cameron Marlow. 2011. Social Capital on Facebook: Differentiating Uses and Users. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 571–580. https://doi.org/10.1145/1978942.1979023

[11] Moira Burke, Cameron Marlow, and Thomas Lento. 2010. Social Network Activity and Social Well-Being. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI '10)*. Association for Computing Machinery, New York, NY, USA, 1909–1912. https://doi.org/10.1145/1753326.1753613

[12] José Campos and Antonio Dias de Figueiredo. 2001. Searching the unsearchable: Inducing serendipitous insights. In *ICCBR 2001. Proceedings of the workshop program at the fourth international conference on case-based reasoning, Vancouver, BC, Canada, 30 July - 2 August 2001*. SSRN, New York, NY, USA, 1–6.

[13] Xiongfei Cao, Xitong Guo, Douglas Vogel, and Xi Zhang. 2016. Exploring the influence of social media on employee work performance. *Internet Research* 26, 2 (2016), 529–545.

[14] Wynne W Chin. 1998. Commentary: Issues and opinion on structural equation modeling. , vii–xvi pages.

[15] Michael Chmielewski and Sarah C Kucker. 2020. An MTurk crisis? Shifts in data quality and the impact on study results. *Social Psychological and Personality Science* 11, 4 (2020), 464–473.

[16] Samantha Copeland. 2019. On serendipity in science: discovery at the intersection of chance and wisdom. *Synthese* 196, 6 (2019), 2385–2406.

[17] Darcy Del Bosque. 2013. Will you be my friend? Social networking in the workplace. *New Library World* 114, 9/10 (2013), 428–442.

[18] Sonja Dreher. 2014. Social media and the world of work. *Corporate Communications: An International Journal* 19, 4 (2014), 344–356.

[19] Siying Du and Steve Gregory. 2016. The Echo Chamber Effect in Twitter: does community polarization increase? In *COMPLEX NETWORKS 2016. Complex Networks & Their Applications V. Studies in Computational Intelligence*, H. Cherifi, S. Gaito, W. Quattrociocchi, and A. Sala (Eds.). Vol. 693. Springer, Cham, 373–378.

[20] Miguel Pina e Cunha, Stewart R Clegg, and Sandro Mendonça. 2010. On serendipity and organizing. *European Management Journal* 28, 5 (2010), 319–330.

[21] Nathan Eagle and Alex Pentland. 2005. Social serendipity: Mobilizing social software. *IEEE Pervasive Computing* 4, 2 (2005), 28–34.

[22] Thomas Erickson and Wendy A. Kellogg. 2000. Social Translucence: An Approach to Designing Systems That Support Social Processes. *ACM Trans. Comput.-Hum. Interact.* 7, 1 (mar 2000), 59–83. https://doi.org/10.1145/344949.345004

[23] David Gefen, Detmar Straub, and Marie-Claude Boudreau. 2000. Structural equation modeling and regression: Guidelines for research practice. *Communications of the association for information systems* 4, 1 (2000), 7.

[24] Samuel D Gosling, Peter J Rentfrow, and William B Swann Jr. 2003. A very brief measure of the Big-Five personality domains. *Journal of Research in personality* 37, 6 (2003), 504–528.

[25] Rex Hartson. 2003. Cognitive, physical, sensory, and functional affordances in interaction design. *Behaviour & information technology* 22, 5 (2003), 315–338.

[26] Jan B Heide and Allen M Weiss. 1995. Vendor consideration and switching behavior for buyers in high-technology markets. *Journal of marketing* 59, 3 (1995), 30–43.

[27] Natali Helberger, Kari Karppinen, and Lucia D'acunto. 2018. Exposure diversity as a design principle for recommender systems. *Information, Communication & Society* 21, 2 (2018), 191–207.

[28] Jonathan L. Herlocker, Joseph A. Konstan, Loren G. Terveen, and John T. Riedl. 2004. Evaluating Collaborative Filtering Recommender Systems. *ACM Trans. Inf. Syst.* 22, 1 (jan 2004), 5–53. https://doi.org/10.1145/963770.963772

[29] David John Hughes, Moss Rowe, Mark Batey, and Andrew Lee. 2012. A tale of two sites: Twitter vs. Facebook and the personality predictors of social media usage. *Computers in Human Behavior* 28, 2 (2012), 561–569.

[30] Mohammad Hossein Jarrahi. 2017. Social Media, Social Capital, and Knowledge Sharing in Enterprise. *IT Professional* 20, 4 (2017), 37–45.

[31] Phillip Jeffrey and Andrew McGrath. 2000. Sharing Serendipity in the Workplace. In *Proceedings of the Third International Conference on Collaborative Virtual Environments* (San Francisco, California, USA) *(CVE '00)*. Association for Computing Machinery, New York, NY, USA, 173–179. https://doi.org/10.1145/351006.351037

[32] Steven Johnson. 2011. *Where good ideas come from: the seven patterns of innovation.* Penguin, London, UK.

[33] Byungkyu Kang, Nava Tintarev, Tobias Höllerer, and John O'Donovan. 2016. What am I not seeing? An interactive approach to social content discovery in microblogs. In *International Conference on Social Informatics*. Springer, Cham, 279–294.

[34] Sara Kiesler and Jonathon N Cummings. 2002. What do we know about proximity and distance in work groups? A legacy of research. *Distributed work* 1 (2002), 57–80.

[35] Bart P Knijnenburg, Martijn C Willemsen, Zeno Gantner, Hakan Soncu, and Chris Newell. 2012. Explaining the user experience of recommender systems. *User Modeling and User-Adapted Interaction* 22, 4-5 (2012), 441–504.

[36] Ansgar Koene, Elvira Perez, Christopher James Carter, Ramona Statache, Svenja Adolphs, Claire O'Malley, Tom Rodden, and Derek McAuley. 2015. Ethics of personalized information filtering. In *International Conference on Internet Science*. Springer, Cham, 123–132.

[37] Jacqueline Lane, Ina Ganguli, Patrick Gaule, Eva Guinan, and Karim R Lakhani. 2019. Engineering Serendipity: The Role of Cognitive Similarity in Knowledge Sharing and Knowledge Production. *Technology & Operations Mgt. Unit Working Paper* 20-058 (2019), 1–75.

[38] Frieder R Lang, Dennis John, Oliver Lüdtke, Jürgen Schupp, and Gert G Wagner. 2011. Short assessment of the Big Five: Robust across survey methods except telephone interviewing. *Behavior research methods* 43, 2 (2011), 548–567.

[39] Christoph Lutz, Miriam Meckel, and Giulia Ranzini. 2013. Trusted Surprises?: Antecedents of Serendipitous Encounters Online. In *63rd Annual Conference of the ICA International Communication Association*. ICA International Communication Association, London, UK, 32.

[40] Christoph Lutz, Christian Pieter Hoffmann, and Miriam Meckel. 2017. Online serendipity: A contextual differentiation of antecedents and outcomes. *Journal of the Association for Information Science and Technology* 68, 7 (2017), 1698–1710. https://doi.org/10.1002/asi.23771 arXiv:https://asistdl.onlinelibrary.wiley.com/doi/pdf/10.1002/asi.23771

[41] S Makri. 2016. Supporting Serendipity in future digital information environments. Teoksessa S. Makri & TM Race (toim.), Accidental Information Discovery (ss. 105–113). London: Chandos.

[42] Stephann Makri, Ann Blandford, Mel Woods, Sarah Sharples, and Deborah Maxwell. 2014. "Making my own luck": Serendipity strategies and how to support them in digital information environments. *Journal of the Association for Information Science and Technology* 65, 11 (2014), 2179–2194.

[43] Kim Martin and Anabel Quan-Haase. 2017. "A process of controlled serendipity": An exploratory study of historians' and digital historians' experiences of serendipity in digital environments. *Proceedings of the Association for Information Science and Technology* 54, 1 (2017), 289–297.

[44] Sabine Matook, Jeff Cummings, and Hillol Bala. 2015. Are you feeling lonely? The impact of relationship characteristics and online social network features on loneliness. *Journal of Management Information Systems* 31, 4 (2015), 278–310.

[45] Abigail McBirnie. 2008. Seeking serendipity: the paradox of control. *Aslib Proceedings: New Information Perspectives* 60, 6 (2008), 600–618.

[46] Lori McCay-Peet and Anabel Quan-Haase. 2016. The Influence of Features and Demographics on the Perception of Twitter as a Serendipitous Environment. In *Proceedings of the 27th ACM Conference on Hypertext and Social Media* (Halifax, Nova Scotia, Canada) *(HT '16)*. Association for Computing Machinery, New York, NY, USA, 333–335. https://doi.org/10.1145/2914586.2914609

[47] Lori McCay-Peet and Elaine G Toms. 2015. Investigating serendipity: How it unfolds and what may influence it. *Journal of the Association for Information Science and Technology* 66, 7 (2015), 1463–1476.

[48] Joanna McGrenere and Wayne Ho. 2000. Affordances: Clarifying and Evolving a Concept. In *Proceedings of Graphics Interface 2000* (Montréal, Québec, Canada) *(GI 2000)*. Canadian Human-Computer Communications Society, Toronto, Ontario, Canada, 179–186. https://doi.org/10.20380/GI2000.24

[49] Alexander Newman, Yuen Lam Bavik, Matthew Mount, and Bo Shao. 2021. Data collection via online platforms: Challenges and recommendations for future research. *Applied Psychology* 70, 3 (2021), 1380–1402.

[50] Alex Newson, Deryck Houghton, and Justin Patten. 2008. *Blogging and other social media: Exploiting the technology and protecting the enterprise*. Gower Publishing, Ltd., UK.

[51] Casey Newton. 2018. Twitter is relaunching the reverse-chronological feed as an option for all users starting today. https://www.theverge.com/2018/12/18/18145089/twitter-latest-tweets-toggle-ranked-feed-timeline-algorithm. Online; accessed 8 April 2020.

[52] Jacqueline Ng, Ina Ganguli, Patrick Gaule, and Karim R Lakhani. 2019. Engineering Serendipity: Atypical Encounters, Collaborations, and Knowledge Production. *Academy of Management Proceedings* 2019, 1 (2019), 15839.

[53] Tien T. Nguyen, Pik-Mai Hui, F. Maxwell Harper, Loren Terveen, and Joseph A. Konstan. 2014. Exploring the Filter Bubble: The Effect of Using Recommender Systems on Content Diversity. In *Proceedings of the 23rd International Conference on World Wide Web* (Seoul, Korea) *(WWW '14)*. Association for Computing Machinery, New York, NY, USA, 677–686. https://doi.org/10.1145/2566486.2568012

[54] Mehrbakhsh Nilashi, Dietmar Jannach, Othman bin Ibrahim, Mohammad Dalvi Esfahani, and Hossein Ahmadi. 2016. Recommendation quality, transparency, and website quality for trust-building in recommendation agents. *Electronic Commerce Research and Applications* 19 (2016), 70–84.

[55] Jennifer E Nutefall and Phyllis Mentzell Ryder. 2010. The serendipitous research process. *The Journal of Academic Librarianship* 36, 3 (2010), 228–234.

[56] Ariane Ollier-Malaterre, Nancy P Rothbard, and Justin M Berg. 2013. When worlds collide in cyberspace: How boundary work in online social networks impacts professional relationships. *Academy of Management Review* 38, 4 (2013), 645–669.

[57] Ekaterina Olshannikova, Thomas Olsson, Jukka Huhtamäki, Susanna Paasovaara, and Hannu Kärkkäinen. 2020. From Chance to Serendipity: Knowledge Workers' Experiences of Serendipitous Social Encounters. *Advances in Human-Computer Interaction* 2020 (2020), 18 pages.

[58] Thomas Olsson, Pradthana Jarusriboonchai, Paweł Woźniak, Susanna Paasovaara, Kaisa Väänänen, and Andrés Lucero. 2020. Technologies for enhancing collocated social interaction: review of design solutions and approaches. *Computer Supported Cooperative Work (CSCW)* 29, 1 (2020), 29–83.

[59] Salvatore Parise, Eoin Whelan, and Steve Todd. 2015. How Twitter users can generate better ideas. *MIT Sloan Management Review* 56, 4 (2015), 21.

[60] Eli Pariser. 2011. *The filter bubble: What the Internet is hiding from you*. Penguin, London, England.

[61] Henri Pirkkalainen, Ekaterina Olshannikova, Thomas Olsson, and Jukka Huhtamäki. 2021. Examining Serendipitous Encounters and Self-Determination in Twitter-Enabled Innovation. *Advances in Human-Computer Interaction* 2021 (2021), 12 pages.

[62] Tammera M Race and Stephann Makri. 2016. *Accidental information discovery: cultivating serendipity in the digital age*. Chandos Publishing, Cambridge, MA.

[63] TS Ragu-Nathan, Monideepa Tarafdar, Bhanu S Ragu-Nathan, and Qiang Tu. 2008. The consequences of technostress for end users in organizations: Conceptual development and empirical validation. *Information systems research* 19, 4 (2008), 417–433.

[64] Matthew K Robison, Katherine I Gath, and Nash Unsworth. 2017. The neurotic wandering mind: An individual differences investigation of neuroticism, mind-wandering, and executive control. *The Quarterly Journal of Experimental Psychology* 70, 4 (2017), 649–663.

[65] J Reynaldo A Santos. 1999. Cronbach's alpha: A tool for assessing the reliability of scales. *Journal of extension* 37, 2 (1999), 1–5.

[66] Gerard Saucier. 1994. Mini-Markers: A brief version of Goldberg's unipolar Big-Five markers. *Journal of personality assessment* 63, 3 (1994), 506–516.

[67] Viviane Sergi and Claudine Bonneau. 2016. Making mundane work visible on social media: a CCO investigation of working out loud on Twitter. *Communication Research and Practice* 2, 3 (2016), 378–406.

[68] Jörgen Skågeby. 2012. The irony of serendipity: disruptions in social information behaviour. *Library hi tech* 30, 2 (2012), 321–334.

[69] Anne Thoring. 2011. Corporate tweeting: Analysing the use of Twitter as a marketing tool by UK trade publishers. *Publishing research quarterly* 27, 2 (2011), 141–158.

[70] Alice Thudt, Uta Hinrichs, and Sheelagh Carpendale. 2012. The Bohemian Bookshelf: Supporting Serendipitous Book Discoveries through Information Visualization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 1461–1470. https://doi.org/10.1145/2207676.2208607

[71] Jeffrey W Treem and Paul M Leonardi. 2013. Social media use in organizations: Exploring the affordances of visibility, editability, persistence, and association. *Annals of the International Communication Association* 36, 1 (2013), 143–189.

[72] Help Center Twitter. 2020. Personalization based on where you see Twitter content across the web. https://help.twitter.com/en/using-twitter/tailored-suggestions. Online; accessed 8 April 2020.

[73] Ward van Zoonen, Joost WM Verhoeven, and Rens Vliegenthart. 2016. How employees use Twitter to talk about work: A typology of work-related tweets. *Computers in Human Behavior* 55 (2016), 329–339.

[74] Saúl Vargas and Pablo Castells. 2011. Rank and Relevance in Novelty and Diversity Metrics for Recommender Systems. In *Proceedings of the Fifth ACM Conference on Recommender Systems* (Chicago, Illinois, USA) *(RecSys '11)*. Association for Computing Machinery, New York, NY, USA, 109–116. https://doi.org/10.1145/2043932.2043955

[75] Dhaval Vyas, Alan Dix, and Gerrit C van der Veer. 2015. Reflections and encounters: Exploring awareness in an academic environment. *Computer Supported Cooperative Work (CSCW)* 24, 4 (2015), 277–317.

[76] Shanshan Zhang, Ron Chi-Wai Kwok, Paul Benjamin Lowry, and Zhiying Liu. 2019. Does more accessibility lead to more disclosure? Exploring the influence of information accessibility on self-disclosure in online social networks. *Information Technology & People* 32, 3 (2019), 754–780.

[77] Dejin Zhao and Mary Beth Rosson. 2009. How and why people Twitter: the role that micro-blogging plays in informal communication at work. In *Proceedings of the ACM 2009 international conference on Supporting group work*. ACM, New York, NY, USA, 243–252.

PUBLICATION
III

Conceptualizing big social data

Olshannikova, E., Olsson, T., Huhtamäki, J., and Kärkkäinen, H.,

Journal of Big Data, *4*(3), 1–19
https://doi.org/10.1186/s40537-017-0063-x

Olshannikova *et al. J Big Data (2017) 4:3*
DOI 10.1186/s40537-017-0063-x

Journal of Big Data

SURVEY PAPER

Open Access

Conceptualizing Big Social Data

Ekaterina Olshannikova[1*], Thomas Olsson[1], Jukka Huhtamäki[2] and Hannu Kärkkäinen[3]

*Correspondence: ekaterina.olshannikova@tut.fi
[1] Department of Pervasive Computing, Tampere University of Technology, Korkeakoulunkatu 10, 33720 Tampere, Finland
Full list of author information is available at the end of the article

Abstract

The popularity of social media and computer-mediated communication has resulted in high-volume and highly semantic data about digital social interactions. This constantly accumulating data has been termed as Big Social Data or Social Big Data, and various visions about how to utilize that have been presented. However, as relatively new concepts, there are no solid and commonly agreed definitions of them. We argue that the emerging research field around these concepts would benefit from understanding about the very substance of the concept and the different viewpoints to it. With our review of earlier research, we highlight various perspectives to this multi-disciplinary field and point out conceptual gaps, the diversity of perspectives and lack of consensus in what Big Social Data means. Based on detailed analysis of related work and earlier conceptualizations, we propose a synthesized definition of the term, as well as outline the types of data that Big Social Data covers. With this, we aim to foster future research activities around this intriguing, yet untapped type of Big Data.

Keywords: Big Social Data, Social Big Data, Digital human, Conceptualization, Social Data, Social media, Computational social science, Social computing, Classification, Big Social Data analysis

Introduction

Background

We live in an "always-on society" [1–3], meaning that people constantly interact with each other. Due to the rapid development of social computing and mushrooming of social media services, much of social interaction is nowadays mediated by information technology and takes place in the digital realm. An average Internet user consumes and shares large amounts of digital content every day through popular social online services, such as Facebook, Twitter, YouTube, Instagram and SnapChat.

From data perspective, this has led to emergence of extensive amounts of human-generated data [4, 5] with diverse social uses and rich meanings (for example, communication text, videos for entertainment and self-representation, sharing of news and other 3rd party content in social media). Such unstructured/semi-structured, yet semantically rich data has been argued to constitute 95% of all Big Data [6]. This Social Data explosion has resulted in theorizations and studies about the emerging topic of Big Social Data (BSD).

Broadly speaking, BSD refers to large data volumes that relate to people or describe their behavior and technology-mediated social interactions in the digital realm. The sheer volume and semantic richness of such data opens enormous possibilities for utilizing and analyzing it for personal [7, 8], commercial [9, 10] as well as societal purposes [11–13]. For example, the scattered social media would benefit from meta-services that bring together all the content from a user. Commercial use could include even more targeted advertising, matchmaking services, or many unimaginable data-centered business models [14, 15]. The search for beneficial applications and services in regard to BSD has only just begun.

Central concepts and goals of the research

In the research literature, the concept of Big Social Data has been defined and interpreted in many ways for various purposes; for example, the viewpoints from which it has been explored include social media, online social networks, social computing, and computational social science (CSS). The role of these fields in the scope of BSD is discussed in detail in the following sections.

As a rule, BSD is mainly utilized to extract insights from social media data and online social interactions of people for descriptive or predictive purposes to influence human decision-making in various application domains [16–18]. In general, researchers have focused on the analytics and utilization, having paid little attention to clarifying the very concept of BSD and understanding the related phenomena (for example, [19–21]).

In fact, there seems to be lack of consensus about the definition of BSD and the related terms, as we will analyze in the upcoming sections. Inconsideration of proper conceptualization may bring researchers methodological challenges in their studies, especially in such inherently broad and multi-disciplinary field as BSD.

Therefore, we argue for conceptual and theoretical work about the concept of BSD in order to inform future research activities as well as to foster the practical utilization of the data, which may signify social insight. There is a timely need to describe, review, and reflect on BSD literature in order to bring clarity to the concept and understanding about its beneficial opportunities for the practitioners of computational social science and other related research fields.

The potential value of this paper for the readers is presented as follows:

1. Firstly, by the literature review we aim to bring clarity on various existing BSD concepts and its definitions. We discuss relations between BSD and related fields of science in order to inform readers about the domains where this concept is currently applied. We consider these aspects will help researchers to properly identify scope and directions for their investigation on the topic;
2. Secondly, by providing a synthesized concept and definition of BSD we want to motivate researchers to develop better conceptualizations and clarifications of the BSD meaning in regard to their research. Currently, the majority of papers related to the topic are focused on analytical tasks and methods missing the explanation about what researchers consider as BSD and why. As an improvement step towards a holistic approach to this emerging field, BSD practitioners can utilize the definition presented in this work by revising it according to their research objectives;
3. By providing a comprehensive list of BSD types we aim to inform researchers about categories of data that is currently available for research and analysis. This serves as a starting point to identify research opportunities and practical means towards

data-driven research. It is worth noting that there is no extensive taxonomy of BSD in related literature and we neither aim to design one; however, our classification of such data serves as an inducement to the research community for collaboratively creating this taxonomy;

4. Moreover, by describing the key characteristics of BSD we differentiate it from the concept of Big Data. By doing so, we anticipate the emphasis on its unique qualities to open new opportunities for multi-disciplinary research ventures.

In general, we assume this work will attract researchers' attention to explore the holistic view on BSD concept and help them to identify relevant sources of data to utilize in BSD studies.

Related concepts and literature

Due to rapid development of online social services and tremendous growth of data therein, various concepts have emerged in different research fields to help understanding digital environments and their social effects. This section reviews related concepts relevant to BSD and their correlations, as well as outlines existing literature on the topic (see Fig. 1).

There are many interpretations and terms to refer to the "social" aspect in Big Data. The most widespread terms so far are Social Big Data (SBD) and Big Social Data (BSD). Various definitions and approaches are presented and compared in the following, in order to outline the existing research directions.

Big Social Data as science: Ishikawa's and Pentland's concepts

Hiroshi Ishikawa is a central adherent of Social Big Data concept, which he described and defined in his book as science of analyzing interconnections between physical world data and social data for the good of public:

"Analyzing both physical real world data (heterogeneous data with implicit semantics such as science data, event data, and transportation data) and social data (social media data with explicit semantics) by relating them to each other, is called Social Big Data science or Social Big Data for short" [22].

It is worth noting that Ishikawa is one among few who provide a proper conceptualization of his ideas and views on the social phenomenon in Big Data. Accordingly, he clarified and supported by arguments relevant related terms, data sources and analytical approaches.

Thus, he defines *social data* as *social media data*, which, in his opinion, is one kind of Big Data with four V's characteristics—*volume*, *variety*, *velocity* and *vague*. While the first three and *veracity* characteristics are already discussed in multiple studies on Big Data [23–26], the *vagueness* first appears in this book as essential characteristic of social data. It should not be mixed with *vagueness* proposed by Venkat Krishnamurthy on Big Data Innovation Summit in Silicon Valley in 2014, which refers to the confusion over the meaning of Big Data [27–29]. According to Ishikawa, vagueness characteristic is a result of a combination of various types of data to be analyzed, which lead to inconsistency and deficiency. It also relates to the issues of privacy and data management as social data involves individuals' personal information.

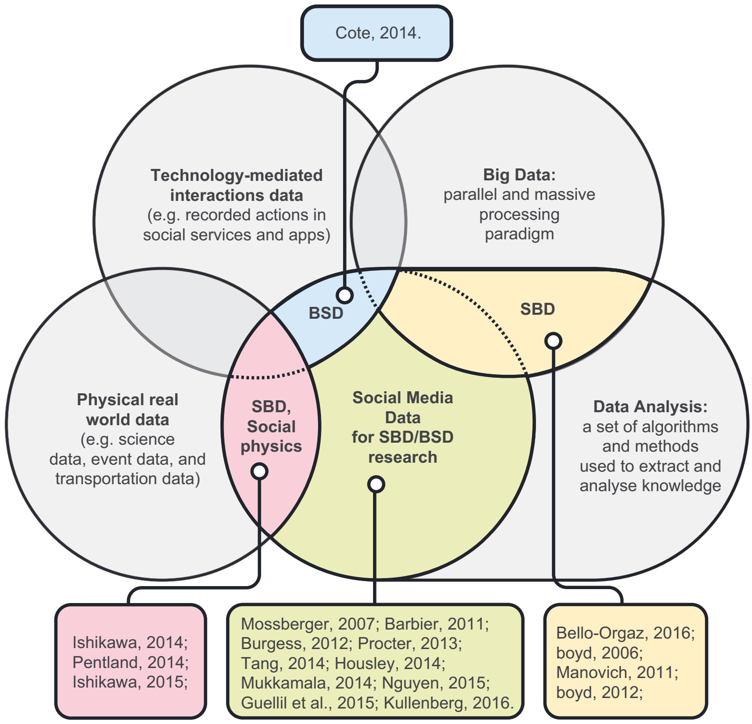


Fig. 1 Conceptual map of various BSD/SBD interpretations in the related literature. This illustration depicts four main domains, which were studied by different researchers from various perspectives and intersections of science field/data types

Additionally, Ishikawa classifies the sources of social media data accordingly: *blogging, micro blogging, social network services, sharing* and *video communication services, social news* and *gaming, social search* and *crowd sourcing services*, and *collaboration services*. All data in such services would therefore be regarded as Big Social Data.

Ishikawa is interested in relationships between physical and cyber worlds. He considers SBD should follow the bidirectional analysis that includes influences from the physical real world on social media, and vice versa, in order to develop a complete model (theory). Such theory may explain interactions between both realms and enable potential prediction, recommendation and problem solving. In other words, he suggests tracking social media data and physical world data in order to reveal mutual interdependencies that in turn would result in actual insight. Ishikawa provides an example of traffic authorities predicting public transportation issues in context of massive social events that are actively discussed in social networks, blogs, news, etc. Thus, the data from social media could be analyzed to prevent traffic jams or to increase the amount of public transportation next to the event location.

Ishikawa's thinking is in line with Pentland's concept of social physics [30]. According to Pentland, social physics is the *"quantitative social science that describes reliable,* *mathematical connections between information and idea flow on the one hand and people's behavior on the other"*. While Ishikawa aims to bring clarity about analytical techniques for SBD (for example, modeling, data mining, multivariate analysis), Pentland envisions a data-driven society. Even though Pentland does not utilize SBD or BSD terms directly in the conceptualization, he defines Big Data as the engine of social physics. The author refers to the data about human behavior, which consists of both human-generated content (from social media platforms) and data from the physical world (for instance, transactions, locations, call records), which is similar to Ishikawa's vision about social data sources. The main goal of Petland's research is to show how this data together with social science theories could be applied in practical settings.

Data-driven approaches to Big Social Data

Guellil and Boukhalfa consider SBD as a part of social computing [31]. To differentiate their view on SBD from general Big Data, authors provide certain characteristics referring to the research of Tang et al. [32]: *"the set of links (due to relationships between users), a nonstructural nature (due to the length of messages required by some microblogging, the presence of spelling mistakes or other) and the lack of completeness (due to certain user requirements for data privacy)"*. Authors provide a classification of the research works on SBD and discuss various analytical approaches and related challenges.

Guellil and Boukhalfa compile their vision of SBD based on the works of Barbier [33], Mukkamala [34] and Nguyen [35]. Notably, Mukkamala and Nguyen utilize SBD and BSD terms interchangeably and mention only social media data as a major data source. Even though Guellil and Boukhalfa point out the inconsistent use of terms in related literature, they do not provide clear conceptualization of the SBD in their own research. In fact, SBD term from the perspective of Guellil and Boukhalfa might be interpreted as a synonym of social media data with qualities such as *large volume*, *noisiness* and *dynamism* that were already revealed earlier in Barbier's work.

From another perspective, Mark Coté makes the attempt to distinguish BSD concept from the broader category of Big Data [36]. In his viewpoint, Big Data is any data produced as the result of the quantification of the world that may include data from sensors, multiple industrial and domestic networks as well as financial markets, whereas BSD *"comes from the mediated communicative practices of our everyday lives, whenever we go online, use our smartphone, use an app or make a purchase."* Moreover, Cote provides reasoning for the importance of BSD. According to him, the concept is not novel, but may significantly affect the media theory. Among those reasons are: the enormous size of data generated by humans that enables endless future analysis; the symbolic nature of social data that is challenging to process even though it is produced in the structured platform spaces; the infrastructure of BSD is very distributed that require scalable computer architecture and network capacity; challenges related to processing, storing, costs and data regulations.

Purpose-driven approaches: Big Social Data for society

Jean Burgess and Axel Bruns discuss Big Data in terms of social media and use the BSD term to refer to this research area [37]. Their vision is based on Manovich's ideology [38], which is focused on bringing the potential of social or cultural data into *humanities* and *social sciences*. Thus, Jean Burgess and Axel Bruns present the BSD concept by mentioning the shift of Big Data towards media, communication, cultural and computational social science, which has led to the wave of research on digital humanities [39–41]. According to Burgess and Bruns, such changes "...provoked in large part by the dramatic quantitative growth and apparently increased cultural importance of social media—hence, "big social data". Their research is aimed to clarify the role of social media in context of the contemporary media ecology with focus on communication, societal events and the nature of human's engagement by applying computational methods towards Twitter archives. Inspired by the Manovich's concept of BSD they trialled the feasibility of research on the phenomenon in order to reveal potential technical, political and epistemological issues. They identified ethical concerns as well as data accessibility, authenticity and reliability challenges. Based on the results, they stated that research on BSD requires the elaboration of mature conceptual models and methodological priorities.

Housley et al. [42] also take a society-oriented view to discuss Big Data. The authors have been conducting observatory research on the opportunities and challenges of open source social media data in the context of social sciences. They seek for the governance and organization improvements through the sense of civil society by means of 'big and broad' social data. According to authors, the term "big and broad" social data refers to three V's (*volume*, *variety*, *velocity*)—already well-known dimensions of related data, which also might be real-time and dynamic. Accordingly, social media could be used to empower people engagement in civil society through a methodological approach to generate sociological insight as proposed in the paper. William Housley et al. characterize digital innovations with qualities such as interaction, participation and "social" that affect complicated relationships between data and analytical capacity, thus enabling participatory infrastructure for public sociology. Consequently, in this regard, the authors point to "citizen social science", which is aimed to assist social scientists by decreasing the challenges of social media data with the help of volunteers among citizens [43]. Such members of public may contribute with research by recording their knowledge, opinions and beliefs, thus connecting the social science academy and society [44, 45].

Big Social Data as method

Bello-Orgaz et al. [46] consider SBD is a combination of Big Data and social media. According to the authors, SBD is needed for analysis of large amount of data from diverse social media sources. They theorize the concept as follows: "Those processes and methods that are designed to provide sensitive and relevant knowledge to any user or company from social media data sources when data sources can be characterized by their different formats and contents, their very large size, and the online or streamed generation of information".

Thus, the conceptual map of SBD from Gema Bello-Orgaz et al. incorporates *Big Data* as processing paradigm, *social media* as the main source of data, and *Data Analysis* as method gaining and analyzing knowledge. Authors revise analytical methodologies for social media as well as new related applications and frameworks.

Summary of the related literature

Even though not all in the above-mentioned papers explicitly use BSD as a term, we consider these works are relevant to the topic. Researchers try to clarify the phenomenon of rapidly growing amount of human-related social data and seek for ways to apply it for the good of the society, data analytics and various fields of science. The key content of the approaches under discussion and theorizations about BSD is summarized in Table 1.

One central commonality among existing research directions is the presence of social media as major data source and orientation towards analytics. The conceptualizations in these scientific articles vary from fundamentally broad (e.g. Ishikawa [22] and Pentland [30]) to vaguely described (e.g., Guelil and Boukhalfa [31]). Additionally, there are only a few attempts to distinguish the concepts from mere Big Data. What is also important, there is lack of clarity regarding the relations between researchers' concepts and related fields: it is hard to outline how other sciences affect the scope of BSD/SBD and directions of studies. Moreover, it is often confusing what data types are considered relevant and valuable for research, and it is hard to understand which data was utilized in the reported research.

We conclude that there are research gaps that researchers of BSD should bridge in order to achieve holistic understanding about the concept of BSD and its characteristics. For example, it is essential to identify the data types that can be explored and studied in this domain. Sophisticated conceptualization and definition of BSD would help researchers build proper methods to process and analyze it. This is essential also because the growth in human-generated data engenders new challenges to solve, requiring novel tools, frameworks and methodological approaches as well as multidisciplinary expertise.

Theoretical foundations of Big Social Data

Based on the literature overview we perceive the concept of BSD as a combination of four fields of science: social computing (including social media and social networks), Big Data science and data analytics as fields that enable and contribute to the existence of the data, and CSS as a field that primarily utilizes the data to gain insight and conduct research (see Fig. 2).

We emphasize that the concept should be understood in an interdisciplinary way in order to open new research avenues. The current and possible roles of each field of science in the context of BSD are discussed in the following.

Social computing

Social computing is a research and application field that integrates social and computational sciences [47]. According to Wang, the theoretical foundations of social computing incorporate Social Psychology, Sociology, Social Network Analysis, Anthropology as well as theories of organization, communication, human–computer interaction and computing theory. In his work, Kling [48] addresses the idea of a mutual interference between communication technologies and society. Therefore, social computing favorably affects both society and technology development: on the one hand enabling smooth socialization and social interactions through various computational systems, and on the other hand, introducing social practices and theories in the development of computational systems and applications. In terms of BSD, *social computing enables services*

Table 1 Summary of BSD-related concepts, types of data they cover as well as challenges and research goals related to the area

Authors	Proposed conceptualization	Data sources	Data challenges/ characteristics	Research goals
Big Social Data as science				
Ishikawa [22]	SBD—science about relationships between physical world data and social media data	Social media data (explicit semantics); physical real world data (implicit semantics)	Volume; variety; velocity; vague	To clarify fundamental conceptualization of SBD and its applications
Pentland [30]	Social physics—quantitative social science about connections between idea/information flow and human behavior	Physical world and social media data that reveal human behavior	Large size; data is ubiquitous and real-time	To conceptualize social physics; To reveal applications of social physics in real-world settings
Data-driven approaches to Big Social Data				
Guelil and Boukhalfa [31]	Concepts from Barbier [33], Mukkamala [34] and Nguyen [35]. SBD is a part of social computing	Social media data	Lack of completeness; large size; dynamic and unstructured	To review research on SBD and classify related literature; to study analytical approaches to SBD
Coté [36]	Concept of data motility under the age of BSD	Data from mediated humans' practices	Volume; symbolic nature of data; distributed infrastructure; lack of regulations	To conceptualize the data motility; to bring conceptual boundaries of BSD
Big Social Data for society				
Burges and Bruns [37]	BSD is social media data	Social media data	Data authenticity, reliability and accessibility issues	To assess feasibility of research in BSD; to reveal potential issues while working with BSD in academia context
Housley et al. [42]	"Big and broad" social data	Social media data	Volume; variety; velocity; real-time; dynamic	To conduct observatory research of "big and broad" social data opportunities and challenges
Big Social Data as method				
Bello-Orgaz et al. [46]	BSD —processes and methods to extract knowledge from social media to users or companies	Social media data	Volume; velocity; variety; value; veracity	To review technologies and applications for processing Big Data from social media

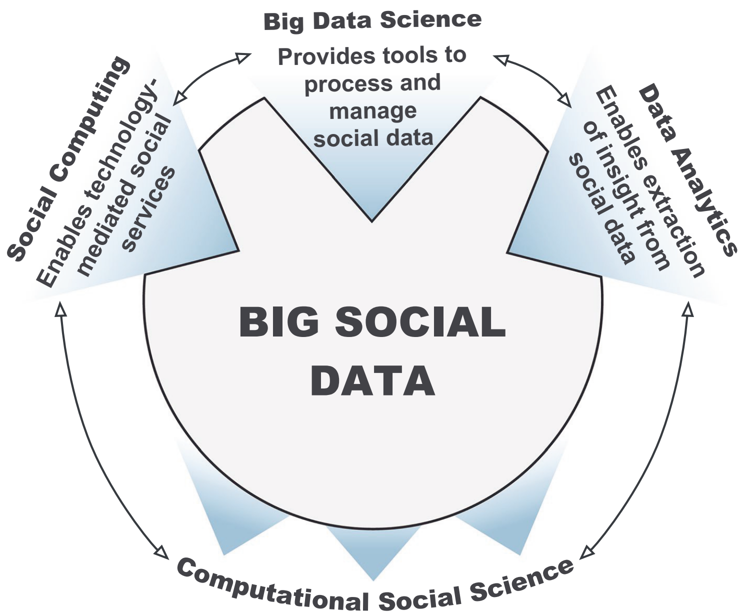


Fig. 2 Different fields of science contributing to and utilizing Big Social Data as a field. There are four main science fields that contribute to and utilize Big Social Data as a research field—social computing, Big Data science, data analytics and CSS

for technology-mediated self-representation [49] *and communication and supports the building and maintaining of digital relationships through multiple technological infrastructures (for example, Web, database, multimedia and wireless technologies).* In summary, social computing approaches the topic from the perspectives of applications, communication and business.

Big data science

Big data science refers to a field that processes and manages *high-volume*, *high-velocity* and *high-variety* data in order to extract reliable and valuable insights [50]. Big Data is aimed to serve large-scale digital applications and computational systems. Therefore, from BSD perspective, *Big Data science provides solutions to process and manage data originated from technology-mediated social interactions in the context of numerous social services and applications in the digital environment.* There are both optimistic and realistic approaches in regard to recent interest to Big Data technology. One group of researchers (as a rule business-oriented) discusses potential benefits of utilizing Big Data [51, 52] to study massive data about people, things and interactions, while other researchers appeal to critical questions, assumptions and issues that may occur when accessing such data [53–55]. It is crucial to consider a critical view on BSD concept, because data that is primarily related to digital human interactions would definitely cause controversial challenges (for example, data availability, regulations on accessing data, ethics issues, and privacy). In summary, originating from computer science and information systems Big Data is a broader category than BSD, and has mostly data and infrastructure-centric perspective, for instance, with focus on Hadoop, Spark, clusters, and related infrastructural work.

Data analytics

Data analytics allows the extraction of insight or conclusions from existing massive data sets. Generally, it includes *descriptive* (describes data), *exploratory* (discovering unknown correlations in data), *predictive* (predict events and trends) and *prescriptive* (suggest actions) methods to gain meaningful insight for different domains [56, 57]. *Social Network Analysis* (SNA) is one of the most established fields of data analytics [58, 59], providing tools, methods and theories for the research of social networks in the digital realm. Other central areas that can be relevant for BSD include *Business Analytics* [60, 61] and *Sentiment Analytics* [62, 63]. Regardless of the intention and application area of the analysis, data analytics can be said to approach BSD from the perspective of utilization of data (for example, service development, gaining insight, decision making).

Computational social science

Definition of the concept is only one step towards proper understanding of BSD. Duncan Watts claimed the potential of Big Data in social domain—"we finally have our telescope" [64]. However, Macy challenges this statement [65] by referring to Gintis and Helbing [66] who point out that just having a telescope is not enough. "We also need to know where to point it, and for that we need the core analytical toolkit... Big data needs big theory" [65]. In terms of BSD such a pointer or a guide toward the theory and meaningful applications is *CSS* [67]. This multidisciplinary field seeks for theory-grounded models of the social phenomena within the intersection of social and computational sciences [68]. CSS determines a joint collaboration between social, behavioral, cognitive and computer scientists with agent theorists, mathematicians and physicists [69]. According to Conte, CSS is going beyond the traditional social science tools to unravel social complexity from new perspectives more deeply [70]. Author highlights that CSS is not only about variables and equations; the major elements of this science are "people, ideas, human-made artifacts, and their relations within ecosystems". The theorization and modeling of society by means of computational approaches is aimed to bring comprehension of social complexity and the way social systems operate [71]. Thus, we argue that CSS utilizes BSD in order to "serve the public good and examine the public agenda" [72]. In other words, CSS can reveal the meaningful and relevant areas in utilization of BSD, thus pointing directions for the analysis, making sense of the findings and enabling predictions as well as sensible explanations.

In summary, the aforementioned areas are the central conceptual and theoretical foundations of BSD that contribute to this inter-disciplinary concept. Social computing enables and serves technology-mediated social services and applications that in turn generate vast amount of complex social data; such data are managed and processed through Big Data tools; then insights and prescriptions are derived from data analytics methods and algorithms. CSS is one of the key fields to define targets and reasons for the analysis and explanations for the analysis results.

Our synthesis and definition of Big Social Data

Drawing from our overview of the related literature and observation of contributing science fields we provide a meta-level definition of the synthesized BSD concept as follows:

Big Social Data is any high-volume, high-velocity, high-variety and/or highly semantic data that is generated from technology-mediated social interactions and actions in digital realm, and which can be collected and analyzed to model social interactions and behavior.

This definition approaches the concept from the synthesized perspective including the description of social data characteristics, its sources and origins as well as purpose of use:

Characteristics Shortly speaking, in this context, *volume* refers to the exponential growth of social data. *Variety* relates to various types and forms of social data sources: it might be structured, semi-structured or unstructured. Variety can also mean the difference of formats (for instance, text, image, video). Velocity refers to the fact that social data is generated and distributed with tremendous speed. One can simply count his/her activity in online services per hour to imagine the frequency, with which billions of people right at this moment create or share something online. These characteristics define the size of social data available for the analysis as well as real-time and dynamic nature of BSD. The volume, velocity and variety are traditional characteristics in any Big Data, while *semantic* is a more unique characteristic of BSD. It refers to the fact that all content manually created is highly symbolic with various often-subjective meanings, which require intelligent solutions to be analyzed. There have been studies on mining and analyzing such multimedia data [73–76], however we are still far from the degree of the intelligence, which may turn immense pools of user-generated content into meaningful insights.

Data sources and origins In context of BSD, we consider technology-mediated social interactions as origins of social data types. It refers to *digital self-representation*, *technology-mediated communication* and *digital relationships data* that may appear not only in social networks services but in variety of discussion forums, blogs, web and mobile chat applications, multi-player games as well as different web sites that are not for social purposes per se.

Purpose Analyzing and modeling social interactions and behavior means that researchers may use the data to describe, understand, and build models of digital interactions taking place between people and how people act (online) around these interactions (for example, profile building, self-expression and other activities that are not directly seen as interaction but, rather, necessary prerequisites for it). The knowledge, which is gained from analysis, may then be utilized in variety of applications, meaning that BSD practitioners are free to choose which domain or research question to address. For instance, researchers may aim to solve fundamental societal issues or just explore tweets for the sake of testing new semantic algorithms.

The definition is further explicated in the following subsection with the classification of data types that relate to technology-mediated social interactions.

Types of Big Social Data

We emphasize that a central element of the BSD concept is "digital human", who uses Information and Communications Technology (ICT) for digital social interactions. The rapid evolution of ICT has shifted the role of a user from a consumer to the active producer and mediator of information, thus allowing people to control, personalize and apply the digital realm according to their values, social needs and preferences [70]. We incorporate the term of "digital human" to underline the shift towards new sociality that lives in hybrid reality [77], where the dynamism and constant availability of technology-mediated communication blurs the boundaries between reality and virtuality. Thus, people do not distinct their activity in online and physical environments, because of "always-on" social networking. Similarly, Wooglar suggests the term of "virtual community" and states that it is just the matter of choosing words: *"In this usage, 'virtual,' like 'interactive,' 'information,' 'global,' 'remote,' 'distance,' 'digital,' 'electronic' (or 'e-'), 'cyber-,' 'network,' 'tele-,' and so on, appears as an epithet applied to various existing activities and social institutions".* [78].

Around digital human interactions, there are both machine-generated and human-generated data that potentially might turn into the social insight. However, in this paper we argue that exactly human-generated data makes BSD concept unique and distinguishes it from general field of Big Data. While machine-generated data could be analyzed through mere Big Data tools and applications, human-generated content requires more intelligent solutions to decode the semantics of people's beliefs, opinions and behavior. Undoubtedly, Big Data may show what and how is changing in social interactions, however it does not answer the question of why those changes and processes are happening. Therefore, we consider BSD is the solution to properly investigate the semantics of human-generated content. From our perspective, it may provide to practitioners of many research fields both facts and reasoning.

While discussing human-generated data we mean content that is produced through social technology-mediated interactions of people in social media platforms. This category may contain digital-self representation data, technology-mediated communication data and digital relationships data (see Fig. 3). These three categories define the types of data that could be interpreted and utilized as social data in the current digital environment (see Table 2). In other words, Table 2 serves as a simplified taxonomy of BSD; however, it is not meant as an extensive index of what data is BSD but, rather, as a list of currently existing BSD examples that could be available for research and analysis.

Digital self-representation

The first category to be discussed is digital self-representation. This is the initial step for "digital humans" to socialize and communicate themselves in the digital realm. These data types relate to numerous virtual profiles that have functions of identity depiction and communicative body [49]. In other words, the data is meant to reveal some information (a "face") for other users in the particular digital service. Albrechstlund proposes a concept of "sharing yourself", which is related to the way constructed identity is participating in social networks creating relations with others [79]. In digital environment people are limited in verbal and non-verbal impressions compensating it by means of text, pictures, videos and music that could be placed in the following data categories:

1. *Profile data* It includes login data (usually a name/nickname/e-mail address with which other people identify the user); identity data (depends on the digital environment, i.e. for some services one should provide real first name and last name, mobile

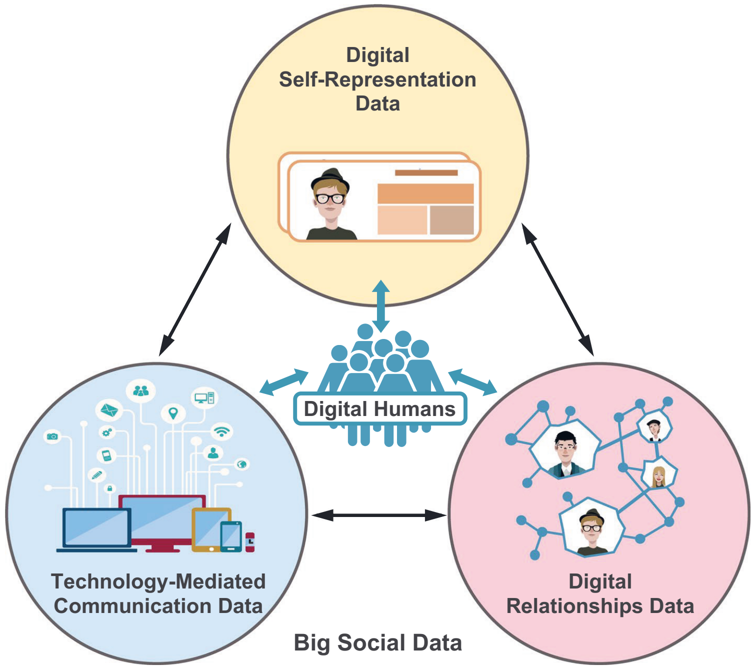


Fig. 3 Overview of BSD types and sources. There are three major data types of Big Social Data—technology-mediated communication data, digital self-representation data and digital relationships data

phone number, country, education, birthday); and personality data (e.g., profile pictures, tags of interest, slogan, personal signature in discussion forums) In many social media services, it is the personality data that the other users particularly focus on and analyze to assess the interestingness of the user.

2. *Self-published content* It incorporates publicly disclosed or socially restricted data (to trusted users or specific communities), such as most status updates in social media, pictures, videos, and other content that people add to services to represent themselves.
3. *Data published by the community* Self-representation could be complemented through person-related content shared by other users. This refers to collaboratively created pictures, narrations, videos, etc.

Technology-mediated communication data

Technology-mediated communication data refers to the data generated in two-way communication, collaborative knowledge creation and information distribution in the context of digital environment—the content and subjects of the communication. Technology mediates the constructed digital self-representation to contribute information, edit existing contributions, comment on entries and discuss related matters. From the fundamental perspective digital environments allow people to contribute to knowledge

Table 2 Classification of Big Social Data types

Category	Definition	Types of data
Digital self-representation data	Data related to *identity depiction* and *communicative body* in digital environment	*Profile data* (i) Login data (name/nickname/e-mail address and password); (ii) Mandatory data (services and application required data, for example, full name, citizenship, birthday); (iii) Extended data (profile pictures, education, tags of interests)
		Self-published content (e.g., personal documents, pictures, videos, interests): (i) Disclosed data (to the public); (ii) Entrusted data (content sharing within trusted digital community)
		Data published by the community (e.g., pictures, narrations, videos, posts): Relates to content shared by other users, which contribute to the digital identity creation
Technology-mediated communication data	Data related to two-way communication, knowledge creation and distribution through technology	*Private communication data* instant 1-to-1 messaging and content sharing;
		Public communication data 1-to-many messaging, commenting, information contribution and editing of existing entries;
		Collaborative communication data many-to-many participatory content sharing, chats, video-conferences
Digital relationships data	Data that reveal digital social relationships patterns	*Explicit data* Friendship data–Followee/Follower data;
		Implicit data Data, which is revealed through technology-mediated communication data (e.g., tweets could be analyzed to infer connections between people)

creation and distribution through various digital devices [80]. Digital environment facilitates physical communication channels resulting in *private communication* (i.e., one-to-one), *public communication* (one-to-many) and *collaborative communication* (many-to-many) data. Depending on the context, public and collaborative communication could also be private within the group of participants, i.e. in case it is a private channel of the organization.

Digital relationships data

Digital Relationships data describes the explicit connections and ties between users in the services. Analysis of this data can reveal social relationship patterns, social network structures and various other sociological and network level phenomena in the digital realm. Digital Representation category firstly contains explicit data, which refers to digital friendships and followership that a user has intentionally and explicitly defined. Technology-mediated social services provide the possibility to build virtual communities based on both physical and online activities (to create networks based on existing connections in physical world and/or create new networks with people from digital realm). There are two roles for users of such services—to be followee and follower. One could have followers or friends on various social platforms (Facebook, LinkedIn, Twitter, Instagram, and many others), and in turn could follow someone to maintain friendships, business relationships or track important content of another relevant user. An interesting factor to be researched is the motivation of people adding someone to the friend's lists. Obviously such lists incorporate friends and colleagues, but also there could be public figures, interesting strangers or people with weak ties [81–83]. There is also implicit data, which could be revealed through analysis of technology-mediated communication data. For instance, tweets can be analyzed to infer individual connections between people. And from these individual connections, we can build network representations of communities in system level. As another example, two users having multiple common contacts (e.g., friend-of-a-friend) can be predicted to become explicit contacts in the future. When a user has, for example, liked or otherwise interacted with a non-contact user's content or profile, there can be seen to be an implicit tie between the users [82]. However, such implicit data normally requires network analysis to be created, and there are few tools or methods to provide such data automatically.

To summarize, we consider this list of BSD types could be valuable for researchers to outline the scope of their interests and will guide them to achieve successful outcomes. Nevertheless, research community has to remember that the accessibility of such data is a crucial challenge of BSD. Lack of access to the data often held by various service providers hinders the utilization of and research opportunities related to this emerging concept. Thus, researches should search for ways of collaboration with social media platforms.

Future work

The holistic overview of related concepts, research fields as well as research communities provide ideas regarding methodological steps that should be taken to enable further research and utilization activities around BSD. This is a combination of three activities that should be primarily focused on in order to open new avenues for the utilization.

1. *Collecting data* The initial step for all researchers who work with BSD is to collect needed datasets for analysis. This step brings up the ethical issues and challenges of data accessibility. Indeed, there are challenges in terms of accessing the data as it is often held by various service providers, which hinders the utilization of the data. Manovich notes this by stating 'only social media companies have access to really large social data' [38]. Fortunately, recently we have seen various movements and joint efforts for bringing together data that, in theory, is public but very challenging to collect in high volume enough for research purposes (for example, the OSoMe[1] project to help analyzing Twitter data). One of the most troubling issues is related to ethics: majority of people are not aware about their data being collected and analyzed by different organizations (including government and social media companies). Moreover, the regulations on accessing and usage of such data are not clear and not completely unified. There are also challenges that may cause privacy violation: collecting more private data than allowed; accessing data without permissions; utilizing data for purposes, which are different from the initial purpose of collecting the data; misinterpreting the data; and changing the content. To make collecting phase feasible we need to fulfill the next step of our framework.
2. *Collaboration* BSD is multidisciplinary area that will require practitioners to build a proper team for work. Our suggestion is to build collaboration with social media platforms or companies that have access to actually large data sets. For instance, the

[1] Observatory on social media (OSoMe) project to study diffusion of information online and discriminate among mechanisms that drive the spread of memes on social media—http://truthy.indiana.edu/about/.

research outcomes from thousands of twits would be questionable in comparison with research under billions of human-generated content from multiple channels. Collaboration with people or companies with various expertize and advantages in terms of social data availability will potentially reduce challenges with collecting data for one's own study, extend the scale and scope of the work in a positive way as well as provide access to multidisciplinary expertise.

3. *Manipulating data* We argue that for gaining meaningful insights from BSD, researchers should design virtual environments where they would be able to access multiple data types, to compare and control them. It may bring new opportunities for authentic and reliable research outcomes. In this regard we agree with Watts [68] that we need *'social supercollider'*, which will obtain diverse social data streams thus opening access to knowledge about people's behavior on the massive scale. BSD artificial environments also could give opportunity to run virtual experiments and validate results with members of related research community.

This paper was aimed to bring clarity on BSD topic in general for any application area. As for our intended future work, we aim to utilize BSD to foster serendipity and, thus, innovativeness in knowledge work organizations. Our objective is to obtain empirical evidence that analysis of BSD can help identify relevant new people to collaborate with.

Conclusion

The multidisciplinary and multi-dimensional nature of Big Social Data brings challenges to the development of a useful conceptualization and definition of the concept. Our literature overview shows that majority of related work on BSD is focused on the analysis of social data, giving less attention to describing what BSD actually is. This can lead to lack of consensus, inconsistency, and vague understanding of what such data could be used for. To bring clarity and sophisticated understanding of BSD we propose a synthesized conceptualization and definition of the concept and this growing field. We reviewed existing literature that demonstrates a variety of applications and approaches to study the phenomena around social data. Based on this we outlined the fields of science that determine the scope of BSD (social computing, Big Data science, data analytics and CSS). We assume the knowledge about the involvement of each field would provide researches with the understanding of the expertise that is demanded for conducting research in this field. Additionally, we proposed the classification of BSD types that, from our perspective, well cover the spectrum of data that BSD consists of. In summary, with this paper, we aim to make researchers more informed about what is BSD, on what data to focus as well as motivate them to elaborate better conceptualization, in order to reach clear desirable research outcomes.

Authors' contributions

EO performed the primary literature review and analysis for this work as well as designed graphics. Manuscript was drafted by EO, TO and JH. EO introduced this topic to other authors and coordinate the work process to complete the manuscript. EO, TO, JH and HK worked together to develop the article's framework and focus. All authors read and approved the final manuscript.

Author details

[1] Department of Pervasive Computing, Tampere University of Technology, Korkeakoulunkatu 10, 33720 Tampere, Finland. [2] Department of Mathematics, Tampere University of Technology, Korkeakoulunkatu 10, 33720 Tampere, Finland. [3] NOVI research group, Department of Information Management and Logistics, Tampere University of Technology, Korkeakoulunkatu 10, 33720 Tampere, Finland.

Acknowledgements
We thank all members of the COBWEB project.

Competing interests
The authors declare that they have no competing interests.

Funding
This work was supported by the Academy of Finland project 295893, 295894, 295895— "Enhancing Knowledge Work and Co-creation with Analysis of Weak Ties in Online Services (COBWEB)".

Received: 1 November 2016 Accepted: 12 January 2017
Published online: 25 January 2017

References

1. Belsey B. Cyberbullying: an emerging threat to the "always on" generation. Recuperado el. 2005; 5. Retrieved from http://www.cyberbullying.ca/pdf/Cyberbullying_Article_by_Bill_Belsey.pdf. Accessed 15 Oct 2016.
2. Katz JE. Handbook of mobile communication studies. London: The MIT Press; 2008.
3. Mandiberg M. The social media reader. New York: NYU Press, New York University; 2012.
4. Monash C. Three broad categories of data. 2010. http://www.dbms2.com/2010/01/17/three-broad-categories-of-data/. Accessed 15 Oct 2016.
5. Chen W. How to tame big bad data. 2010. http://blog.magnitudesoftware.com/2010/08/25/tame-big-bad-data/. Accessed 15 Oct 2016.
6. Gandomi A, Haider M. Beyond the hype: big data concepts, methods, and analytics. Int J Inf Manag. 2015;35(2):137–44.
7. Marwick AE. Status update: celebrity, publicity, and branding in the social media age. New Haven, USA: Yale University Press; 2015.
8. Freire FC. Online digital social tools for professional self-promotion. A state of the art review. Revista Latina de Comunicación Social. 2015;70:288–99.
9. Shih C. The facebook era: tapping online social networks to build better products, reach new audiences, and sell more stuff. Upper Saddle River: Prentice Hall; 2009.
10. Stephen AT, Toubia O. Deriving value from social commerce networks. J Mark Res. 2010;47(2):215–28.
11. Musacchio M, Panizzon R, Zhang X, Zorzi V. A linguistically-driven methodology for detecting impending disasters and un-folding emergencies from social media messages. In: proceedings of LREC 2016 workshop. EMOT: emotions, metaphors, ontology and terminology during disasters; 2016. p. 26–33.
12. Aradau C, Blanke T. Politics of prediction: security and the time/space of governmentality in the age of big data. European Journal of Social Theory. 2016:1–19. Retrieved from http://journals.sagepub.com/doi/abs/10.1177/1368431016667623. Accessed 15 Oct 2016.
13. Saldana-Perez AMM, Moreno-Ibarra M. Traffic analysis based on short texts from social media. Int J Knowl Soc Res. 2016;7(1):63–79.
14. Qualman E. Socialnomics: how social media transforms the way we live and do business. Hoboken: Wiley; 2010.
15. Kennedy H. Commercial mediations of social media data. London: Springer; 2016. p. 99–127.
16. Golbeck J, Robles C, Turner K. Predicting personality with social media. In: CHI'11 Extended abstracts on human factors in computing systems. Vancouver: ACM; 2011. p. 253–62.
17. Power DJ, Phillips-Wren G. Impact of social media and Web 2.0 on decision-making. J Decis Syst. 2011;20(3):249–61.
18. Golbeck J. Big social data predicting the future of you. Executive Tallent Mag. 2014;5:12–4.
19. Cambria E, Rajagopal D, Olsher D, Das D. Big social data analysis. In: Akerkar R, editor. Big Data Computing. Boca Raton, Florida: Chapman and Hall/CRC; 2013. p. 401–14.
20. Bravo-Marquez F, Mendoza M, Poblete B. Meta-level sentiment models for big social data analysis. Knowl Based Syst. 2014;69:86–99.
21. Pandarachalil R, Sendhilkumar S, Mahalakshmi G. Twitter sentiment analysis for large-scale data: an unsupervised approach. Cogn Comput. 2015;7(2):254–62.
22. Ishikawa H. Social big data mining. Boca Raton: Taylor & Francis Group, CRC Press; 2015.
23. Sicular S. Gartner's big data definition consists of three parts, not to be confused with three "V's", vol. 27. Stanford: Gartner, Inc; 2013.
24. Kaisler S, Armour F, Espinosa JA, Money W. Big data: issues and challenges moving forward. In: 2013 46th Hawaii international conference on system sciences (HICSS). New York: IEEE; 2013. p. 995–1004.
25. Tole AA, et al. Big data challenges. Database Syst J. 2013;4(3):31–40.
26. Chen M, Mao S, Zhang Y, Leung VC. Big data: related technologies, challenges and future prospects. In: Springerbriefs in computer science. Cham: Springer; 2014.
27. Borne K. Top 10 big data challenges—a serious Look at 10 big data V's. 2014. https://www.mapr.com/blog/top-10-big-data-challenges-serious-look-10-big-data-vs. Accessed 15 Oct 2016.
28. Kehoe M. What does it take to qualify as 'big data'? 2014. http://www.enterprisesearchblog.com/2014/07/hadoop-salvation-or-hype.html. Accessed 15 Oct 2016.
29. Moorthy J, Lahiri R, Biswas N, Sanyal D, Ranjan J, Nanath K, Ghosh P. Big data: prospects and challenges. J Decis Makers. 2015;40:74–96.
30. Pentland A. Social physics: how good ideas spread-the lessons from a new science. New York: The Penguin Press, Penguin Group; 2014.

31. Guellil I, Boukhalfa K. Social big data mining: a survey focused on opinion mining and sentiments analysis. In: 2015 12th international symposium on programming and systems (ISPS). New York: IEEE; 2015. p. 1–10.
32. Tang J, Chang Y, Liu H. Mining social media with social theories: a survey. ACM SIGKDD Explor Newsl. 2014;15(2):20–9.
33. Barbier G, Liu H. Data mining in social media. Berlin: Springer; 2011. p. 327–52.
34. Mukkamala RR, Hussain A, Vatrapu R. Fuzzy-set based sentiment analysis of big social data. In: Enterprise distributed object computing conference (EDOC), 2014 IEEE 18th international. New York: IEEE; 2014. p. 71–80.
35. Nguyen DT, Hwang D, Jung JJ. Time–frequency social data analytics for understanding social big data. Cham: Springer; 2015.
36. Coté M. Data motility: the materiality of big social data. Cult Stud Rev. 2014;20(1):121.
37. Burgess J, Bruns A. Twitter archives and the challenges of "big social data" for media and communication research. M/C J. 2012;15(5):1–7.
38. Manovich L. Trending: the promises and the challenges of big social data. Debates Digit Humanit. 2011;2:460–75.
39. Berry D. Understanding digital humanities. London: Palgrave Macmillan, Springer Nature; 2012.
40. Kaplan F. A map for big data research in digital humanities. Front Digit Humanit. 2015;2:1.
41. Svensson P. Big digital humanities: imagining a meeting place for the humanities and the digital. Ann Arbor: University of Michigan Press; 2016.
42. Housley W, Procter R, Edwards A, Burnap P, Williams M, Sloan L, Rana O, Morgan J, Voss A, Greenhill A. Big and broad social data and the sociological imagination: a collaborative response. Big Data Soc. 2014;1(2):2053951714545135.
43. Procter R, Housley W, Williams M, Edwards A, Burnap P, Morgan J, Rana O, Klein E, Taylor M, Voss A, Choi C, Mavros P, Hudson Smith A, Thelwall M, Ferne T, greenhill A. Enabling social media research through citizen social science. In: Korn M, Colomnbino T, Lewkowicz M (eds) ECSCW 2013 Adjunct Proceedings, 13th european conference on computer supported cooperative work, 21–25 September 2013, Paphos, Cyprus
44. Mossberger K, Tolbert CJ, McNeal RS. Digital citizenship: the internet, society, and participation. London: MIt Press; 2007.
45. Kullenberg C, Kasperowski D. What is citizen science? A scientometric meta-analysis. PLoS One. 2016;11(1):0147152.
46. Bello-Orgaz G, Jung JJ, Camacho D. Social big data: recent achievements and new challenges. Inf Fusion. 2016;28:45–59.
47. Wang F-Y, Carley KM, Zeng D, Mao W. Social computing: from social informatics to social intelligence. IEEE Intell Syst. 2007;22(2):79–83.
48. Kling R. What is social informatics and why does it matter? Inf Soc. 2007;23(4):205–20.
49. Boyd D, Heer J. Profiles as conversation: networked identity performance on friendster. In: Proceedings of the 39th annual Hawaii international conference on system sciences (HICSS'06), vol. 3. New York: IEEE; 2006. p. 59.
50. Demchenko Y, De Laat C, Membrey P. Defining architecture components of the big data ecosystem. In: 2014 international conference on collaboration technologies and systems (CTS). New York: IEEE. 2014. p. 104–12.
51. Beyer MA, Laney D. The importance of 'big data': a definition. Stamford: Gartner; 2012. p. 2014–8.
52. James M, Michael C, Brad B, Jacques B, Richard D, Charles R, Angela H. Big data: the next frontier for innovation, competition, and productivity. New York: The McKinsey Global Institute; 2011.
53. Boyd D, Crawford K. Critical questions for big data: provocations for a cultural, technological, and scholarly phenomenon. Inf Commun Soc. 2012;15(5):662–79.
54. Akerkar R. Big data computing. Boca Raton: CRC Press, Taylor & Francis Group; 2013.
55. Vis F. A critical reflection on big data: considering APIs, researchers and tools as data makers. First Monday. 2013;18(10). Retrieved from http://ojs-prod-lib.cc.uic.edu/ojs/index.php/fm/article/view/4878/3755. Accessed 15 Oct 2016.
56. Davenport T. Analytics 3.0: In the new era, big data will power consumers products and services. Brighton, MA: Harvard Business Review. Retrieved from https://hbr.org/2013/12/analytics-30. 2013. Accessed 15 Oct 2016.
57. Bendoly E. Fit, bias, and enacted sensemaking in data visualization: frameworks for continuous development in operations and supply chain management analytics. J Bus Logist. 2016;37(1):6–17.
58. Wasserman S, Faust K. Social network analysis: methods and applications, vol. 8. Cambridge: Cambridge University Press; 1994.
59. Easley D, Kleinberg J. Networks, crowds, and markets: reasoning about a highly connected world. Cambridge: Cambridge University Press, University of Cambridge; 2010.
60. Phillips-Wren G, Iyer LS, Kulkarni U, Ariyachandra T. Business analytics in the context of big data. Commun Assoc Inf Syst. 2015;37:448–72.
61. Duan L, Xiong Y. Big data analytics and business analytics. J Manag Anal. 2015;2(1):1–21.
62. Chen C, Chen F, Cao D, Ji R. A cross-media sentiment analytics platform for microblog. In: Proceedings of the 23rd ACM international conference on multimedia. New York City: ACM; 2015. p. 767–9.
63. Boumaiza AD. A survey on sentiment analysis and visualization. In: Qatar foundation annual research conference proceedings, vol. 2016. Doha: HBKU Press Qatar; 2016. p. 1203.
64. Watts DJ. Everything is obvious: how common sense fails us. New York: Crown Business, Crown Publishing group; 2011.
65. Macy MW, et al. Big theory: a trojan horse for economics? Rev Behav Econ. 2015;2(1–2):161–6.
66. Gintis H, Helbing D, Durkheim E, King ML, Smith A. Homo socialis: an analytical core for sociological theory. Rev Behav Econ. 2015;2(1–2):1–59.
67. Lazer D, Friedman A. The network structure of exploration and exploitation. Adm Sci Q. 2007;52(4):667–94.
68. Watts DJ. Computational social science: exciting progress and future directions. Bridge Front Eng. 2013;43(4):5–10.
69. Wallach H. Computational social science: Toward a collaborative future. In: Alvarez RM, editor. Computational social science: Discovery and prediction. USA: Cambridge Universisty Press; 2016. p. 307–16.
70. Conte R, Gilbert N, Bonelli G, Cioffi-Revilla C, Deffuant G, Kertesz J, Loreto V, Moat S, Nadal J-P, Sanchez A, et al. Manifesto of computational social science. Eur Phys J Spec Top. 2012;214(1):325–46.
71. Cioffi-Revilla C. Introduction to computational social science: principles and applications. London: Springer; 2013.

72. Shah DV, Cappella JN, Neuman WR. Big data, digital media, and computational social science possibilities and perils. Ann Am Acad Political Soc Sci. 2015;659(1):6–13.
73. Zhu X, Wu X, Elmagarmid AK, Feng Z, Wu L. Video data mining: semantic indexing and event detection from the association perspective. IEEE Trans Knowl Data Eng. 2005;17(5):665–77.
74. Wu P, Hoi SCH, Zhao P, He Y. Mining social images with distance metric learning for automated image tagging. In: Proceedings of the fourth ACM international conference on web search and data mining. New York City: ACM; 2011. p. 197–206.
75. Hu X, Liu H. Text analytics in social media. New York: Springer; 2012. p. 385–414.
76. Naaman M. Social multimedia: highlighting opportunities for search and mining of multimedia data in social media applications. Multimed Tools Appl. 2012;56(1):9–34.
77. e Silva ADS. From cyber to hybrid mobile technologies as interfaces of hybrid spaces. Space Cult. 2006;9(3):261–78.
78. Woolgar S. Virtual society? Technology, cyberbole reality. New York: Oxford University Press; 2002.
79. Albrechtslund A. Online social networking as participatory surveillance. First Monday. 2008;13(3). Retrieved from http://firstmonday.org/ojs/index.php/fm/article/view/2142/1949.. Accessed 15 Oct 2016.
80. Ruppert E, Law J, Savage M. Reassembling social science methods: the challenge of digital devices. Theory Cult Soc. 2013;30(4):22–46.
81. Granovetter MS. The strength of weak ties. Am J Sociology. 1973;78(6):1360–80.
82. Gilbert E, Karahalios K. Predicting tie strength with social media. In: Proceedings of the SIGCHI conference on human factors in computing systems. New York City: ACM; 2009. p. 211–20.
83. Haythornthwaite C. Strong, weak, and latent ties and the impact of new media. Inf Soc. 2002;18(5):385–401.

PUBLICATION
IV

Scholars' perceptions of relevance in bibliography-based people recommender system

Olshannikova, E., Olsson, T., Huhtamäki, J., and Yao, P.,

Computer Supported Cooperative Work (CSCW), *28*(3–4), 357–389
https://doi.org/10.1007/s10606-019-09349-w

Computer Supported Cooperative Work (CSCW) (2019) 28:357–389
https://doi.org/10.1007/s10606-019-09349-w

Scholars' Perceptions of Relevance in Bibliography-Based People Recommender System

Ekaterina Olshannikova[1,*], Thomas Olsson[1], Jukka Huhtamäki[1] & Peng Yao[1]
[1]*Tampere University, Tampere, Finland (E-mail: ekaterina.olshannikova@tuni.fi; E-mail: thomas.olsson@tuni.fi; E-mail: jukka.huhtamaki@tuni.fi; E-mail: peng.yao@tuni.fi)*

Abstract. Collaboration and social networking are increasingly important for academics, yet identifying relevant collaborators requires remarkable effort. While there are various networking services optimized for seeking similarities between the users, the scholarly motive of producing new knowledge calls for assistance in identifying people with complementary qualities. However, there is little empirical understanding of how academics perceive relevance, complementarity, and diversity of individuals in their profession and how these concepts can be optimally embedded in social matching systems. This paper aims to support the development of diversity-enhancing people recommender systems by exploring senior researchers' perceptions of recommended other scholars at different levels on a similar–different continuum. To conduct the study, we built a recommender system based on topic modeling of scholars' publications in the DBLP computer science bibliography. A study of 18 senior researchers comprised a controlled experiment and semi-structured interviewing, focusing on their subjective perceptions regarding relevance, similarity, and familiarity of the given recommendations, as well as participants' readiness to interact with the recommended people. The study implies that the homophily bias (behavioral tendency to select similar others) is strong despite the recognized need for complementarity. While the experiment indicated consistent and significant differences between the perceived relevance of most similar vs. other levels, the interview results imply that the evaluation of the relevance of people recommendations is complex and multifaceted. Despite the inherent bias in selection, the participants could identify highly interesting collaboration opportunities on all levels of similarity.

Key words: People recommender systems, Social matching applications, Expert search, Homophily, Diversity, Research collaboration, Social networking, Bibliography data analysis

1. Introduction

In scholarly work, collaboration has become a normative form of knowledge production. Researchers across the social sciences broadly concur that collaboration is the best path to solving complex problems and achieving exceptional results (Frydlinger et al. 2013). Collaboration is promoted as a means of cultivating quality, enhanced resource utilization, and high impact (Hsiehchen et al. 2015). In science and patenting, a substantial shift toward collective work has been found across scientific disciplines and business domains (Wuchty et al. 2007; Börner et al. 2010). In academic research, collaboration takes place on a dyadic level between individuals, amongst research teams, as well as within international consortia. However, identifying new suitable candidates for academic collaboration requires high investment in social networking, and the disciplinary structures can prevent unexpected combinations of individuals.

Following this trend of the increasing importance of collaboration, supporting social networking and encouraging new social encounters have become central design goals in the HCI & CSCW communities. Prior research on so-called social matching (Terveen and McDonald 2005) has particularly looked into people recommender systems (Tsai and Brusilovsky 2016; Guy and Pizzato 2016) and opportunistic matching applications (Mayer et al. 2015a; Mayer et al. 2016) that aim to enable identification of new relevant connections, some of them employing playful approaches and gamification (Paasovaara et al. 2016). There are also prototypes of people recommender systems that specifically aim to match scholars: for instance, expert finding systems (Vassileva et al. 2003; Beham et al. 2010), or event-based mobile applications like *'Find & Connect'* developed by Chin et al. (2014) and experimented at the UbiComp 2011 conference. Considering the rich publication data available in online repositories, prior research has looked into bibliography analysis methods for recommender systems, e.g., DBLP[1]-based systems for researchers (Zaiane et al. 2007).

However, the majority of professional matching systems tend to utilize a similarity-maximizing approach, providing recommendations of like-minded others with similar interests. In this regard, Yuan and Gay (2006) deliberate that homogeneity produces both positive and negative effects on interpersonal communication, community formation, and knowledge work – *"homophily not only unifies, it also divides a network"*. On the one hand, collaboration within a group of people with shared interests can contribute to a safe and trustworthy work environment, enabling cohesive team spirit and ease of communication. On the other hand, it has been found that researching and cooperating with diverse individuals is essential in tasks that aim to create new knowledge (Mollica et al. 2003). Prior work emphasizes that an insightful dialogue between diverse actors can build social capital (Burt 2017) by increasing awareness about external knowledge groups and bridging polarized intellectual communities towards abounding knowledge sharing and idea creation (Argote and Ophir 2002). While research on diversifying item recommendations (Adamopoulos and Tuzhilin 2015; Castells et al. 2015) is gaining interest, few attempts have been made to match people based on diversity (Rajagopal et al. 2017).

Additionally, the evaluation of people recommender systems and matching applications is geared toward the assessment of algorithm effectiveness, with little focus on user perceptions. Although there is well-established research on user-centered evaluation of content recommender systems (Knijnenburg et al. 2012;

Pu et al. 2011), the choice of potential collaborators is significantly different and, therefore, requires contextually operationalized evaluation metrics. Considering the diverse needs of scholars, it is essential to pay attention to subjective perceptions regarding the recommended people and carefully conceptualize factors such as perceived relevance and willingness to follow-up on the recommendations.

To enable the gathering of such data, we developed a simple DBLP-based people recommender system that provides the user with recommendations of other scholars from three different levels of similarity regarding their publication history – low, moderate and high. With a user study that combines a controlled experiment and semi-structured interviewing, we address the following research questions: **(RQ1)** *What level of measured similarity of publication history is preferred in recommendations of potential collaborators?* **(RQ2)** *What specific needs and expectations scholars have in regard to seeking professional collaboration?*

The findings reveal an intriguing mismatch between scholars' intuitive behavior and deliberate intentions regarding potential academic collaboration. While the quantitative results demonstrate participants' general preference to most similar recommended people, the interview data brings up a variety of scholars' needs for connecting with cross-disciplinary and diverse people. Thus, the nature of the collaboration task might influence the perceived relevance of potential candidates, for example, regarding the complementarity of professional roles, skills, and expertise.

The contribution of this work is two-fold: (i) providing empirical findings on subjective perceptions of people recommendation relevance in the context of potential academic partnering, and (ii) presenting the qualitative account of academics' needs in collaboration and factors that might affect their decision in choosing partners. Furthermore, as a methodological contribution, we operationalize measures for subjective opinions on people recommendations in the context of professional academic collaboration.

2. Related work

While various disciplines have studied scientific collaboration in different ways, a general consensus is that collaboration is imperative for effective knowledge production. Bozeman et al. (2013) note that research collaboration is often limited to the notion of co-authorship, and criticize the assumption that cooperation is undoubtedly resulting in a knowledge product (e.g., scientific paper). In this article, we approach scholarly partnering practices, which are beyond co-authorship, and do not require an explicit valuable outcome. In fact, Bozeman and Corley (2004) define key motives for scientific collaboration and propose that the selection of collaborators can be driven by: (i) work ethic attribution and schedule compliance; (ii) shared nationality; (iii) need to mentor junior researchers; (iv) administration request or high reputation; (v) preceding collaboration experience, its quality and personality chemistry; (vi) complementarity of skills. Considering this breadth, in this article we adopt the broad definition by Bozeman et al.

(2013): *"collaboration is a social process whereby human beings pool their human capital for the objective of producing knowledge."*

To emphasize novelty of our contribution and the research gap we cover the following topics. First, we discuss research on supporting social interaction and collaboration in the general context of conferences and introduce works on bibliography-based recommender systems. Then, we deliberate on the concepts of similarity and diversity. Finally, we provide an overview of existing user-centered evaluation metrics in recommender systems.

2.1. Computational support for matching scholars

Supporting experts finding is one of the crucial CSCW design goals to facilitate collaborative knowledge creation and dissemination (Ackerman and McDonald 1996). Over the last two decades, systems for supporting the conference experience have expanded from increasing people's awareness of necessary information at the venue (e.g., schedule and contents) to facilitating social encounters, for example arranging meetings (Nishibe et al. 1998) and general enhancement of attendee interaction with the environment and other people (Dey et al. 1999). One reasonably common approach relates to location and proximity-based services for finding relevant connections (Kawakita et al. 2004; Cox et al. 2003). For instance, *'Find & Connect'*, a social networking mobile application developed by Chin et al. (2014) and experimented during the UbiComp 2011 conference, aims to provide the users with social recommendations based on physical proximity and similarity of interests. The results of the work reveal that users preferred acting on familiar recommended people or friends-of-friends, as well as those who have similar research interests.

There are also generic, platform-like services designed to help the attendees meet new people with shared interests (Zenk et al. 2014). An example of such tool is *'Confer'* (Zhang et al. 2016) which has been tested and deployed in several HCI conferences. *'Conference Navigator'* is another example and has gradually acquired new features and functions (Farzan and Brusilovsky 2007; Wongchokprasitti et al. 2010; Parra et al. 2012). For instance, the most recent version (Brusilovsky et al. 2017) – *'Conference Navigator 3'* – is a community-based recommender system that by utilizing content-based and tag-based analysis methods provide the user with personalized suggestions about people and contents of a conference. The user can explore their research community through interactive social network visualization and connect with similar experts either before or during an event.

Another recent research proposes so-called *'Adaptive Conference Companion'* (Arens-Volland and Naudet 2016) – a mobile application that aims to deliver personalized guidance for attendees of academic events. In addition to utilizing conference data and explicit user input, authors enhance profiling and matchmaking mechanism by extracting bibliographic database (DBLP, GoogleScholar) and social media channels (LinkedIn, ResearchGate and MyScienceWork). They applied a term frequency-inverse document frequency (TF-IDF)[2] algorithm (Beel et al. 2016) for recommending most similar people and sessions in the scope of users' interests. The authors speculate that the majority of participants in the experiment were well-prepared for the conference and already had their schedule which matched with recommendation predictions of relevant content. Unfortunately, the authors did not discuss the effectiveness of a system from the social networking perspective, i.e., how people reacted to and perceived the relevance of people recommendations.

Another vein of research relates to recommending academic collaborations by utilizing bibliography data and social networks analysis and, thus, suggesting candidates with similar research interests. For instance, Kong et al. (2016) utilized topic clustering model to retrieve academic domains, calculate authors' features, and analyze academic collaboration networks. Another group of researchers (Li et al. 2014) explored co-authorship networks to identify relevant collaborators with an already existing academic tie. They observed that scholars' relationships are more complicated in a real-world setting and suggested that future work should go beyond existing co-authorship networks and consider matching people without established connections. Most recently, Hoang et al. (2017) proposes a new approach to calculate similarity with deep learning and experiment on DBLP and WiKiCFP databases. Sie et al. (2012) designed a system for recommending future co-authors that utilizes co-authorship network and topic similarity aspects in the matching mechanisms. They tested researchers preferences regarding existing co-authors vs. new potential candidates with a light-weight user study by asking the participants to rate each recommendation on a scale from one to ten. The findings revealed that participants prefer existing connections with whom collaboration has been already established.

To summarize, the prior work displays a diversity of research and development of people recommender services for scholars with a particular focus on analyzing human-generated content and publishing history. The literature on algorithmic approaches indicates that bibliographical data sets can serve as a valid data source for identifying and recommending social connections. The algorithmic choices encouraged us to approach the area with specific topic modeling methods (TF-IDF) and to analyze the cosine distance[3] (Li and Han 2013) between the authors. Although there is apparent interest in creating services for academic partnering, the primary contribution of preceding research lies in the design of new matching mechanisms and algorithms. The evaluation of such systems, thus, is focused on testing the quality criterion of prediction accuracy and little attention is paid to the subjective perceptions of recommendation usefulness and user's intention to follow-up on those. In this article, we experiment on content-based similarity-difference dimensions and specifically focus on the human-centered and subjective evaluation of perceived relevance and related variables.

2.2. Concepts of similarity and diversity

Our approach of diversity-enhancing people recommender systems is founded on the relatively strong consensus that fruitful collaboration and high innovation capability result from complementary viewpoints among a diverse group of actors (Mitchell and Nicholas 2006). Rodan and Galunic (2004) imply that heterogeneous knowledge is of high importance to both overall managerial performance and particularly to innovation performance. However, the factual value of diversity and how it should exactly manifest remain unclear. Despite the extensive literature, the role of both similarity and diversity (as opposite ends of a continuum), particularly in the decision-making of choosing academic collaborators, requires more research, as will be shown in what follows.

The related work discussed in the previous subsection demonstrates a tendency of utilizing similarity-maximizing approaches for recommending content and connections, thus amplifying the effects of homophily bias. The concept of *homophily* has caught the attention of researchers, primarily in social psychology (Lazarsfeld and Merton 1954; Marsden 1987; Moody 2001), as the phenomenon of individuals' natural preference to interact with similar-minded people who share socio-cultural traits. In CSCW and HCI research, homophily has been addressed, for instance, as a predictive and influential factor of online behavior in content preferences (Chang et al. 2014), and audience attraction on social media (Sharma and Cosley 2016). Another vein of research focuses on studying diversity in terms of human, relational and intellectual capital within global organizations to design features that support online communities in collaborative tasks (Muller et al. 2012). Researchers and developers seem to have adopted the similarity-maximizing approach from item recommender systems, using metrics of similarity as the proxy for relevance also in matching peers within organizations (Guy et al. 2010) and scholars in the context of academic collaboration (Heck 2013).

Although homophily might strengthen existing communities, it does not encourage the creation of new ties to further away in the global social network. Some researchers propose that such mechanisms directly lead to the formation of echo chambers that are detrimental to information flow, innovation, and creativity (Jasny et al. 2015; Bessi 2016). Echo chambers have received critique particularly with respect to social media services that divide the user community into camps of different opinions and thus increase polarization in the society (Li et al. 2013; Lee et al. 2014).

At the same time, organizational studies have identified that also diversity can have negative influences on collaborative activities, such as information exchange and decision-making (Graves and Elsass 2005; Hobman et al. 2004). An extensive review (Mannix and Neale 2005) concludes that social differences (i.e., surface-level), such as race and gender, indeed tend to have adverse effects on the ability of groups to function effectively, whereas more profound cognitive dissimilarities, such as differences in expertise or personality, are more often positively related to team performance. In other words, diversity is strongly linked to the concept of identity, which can make the introduction of diversity challenging in established work cultures.

Following the above mentioned, CSCW research has investigated whether it is possible to overcome the adverse effects of dissimilarities in teams to provoke creativity and productivity. For instance, Dong et al. (2016) found that commitment to a common cause, such as shared goals of the work, bring people together despite cultural differences. Similarly, but from a broader perspective, Ye and Robert Jr. (2017) revealed that collectivism (over individualism) makes people more tolerant to differences in terms of personal values, working styles, skills, and general abilities, thus, embracing individual creativity and work satisfaction. Besides, Rajagopal et al. (2017) investigated how to match peers with dissimilar opinions. The findings demonstrated that matching people with different interpretations of shared interests is more effective in producing positive experiences of breakdown.

Overall, the literature on diversity and homophily contains interesting contradictions, which calls for further empirical research on various forms of similarity or diversity in different types of collaboration. To this end, we seek to uncover how the two concepts are interlinked particularly in the assessment of the relevance of potential scholarly collaborators.

2.3. User-centered evaluation criteria for recommender systems

Historically, research on recommender system has primarily focused on the design of algorithms, underlying the assumption that better algorithms results in better user experience with the systems. Pu et al. (2012) challenges this premise by providing conceptual observation and guidelines on the evaluation criteria for recommender systems. They explicitly emphasize the importance of the user's perception regarding the system qualities. We summarize existing conceptualizations of the recommendation quality as follows: **(i)** *perceived accuracy* (Pu et al. 2011) – how well recommendations match with users interests defines the trust towards the systems; **(ii)** *familiarity* (Sinha and Swearingen 2002) – presence of familiar items increase trust towards the system; **(iii)** *novelty* (Castells et al. 2015) – unexpectedness of received recommendations can affect perceived usefulness of the system; **(iv)** *diversity* (Nguyen et al. 2014) – receiving diverse items lessens filter bubble thus increasing users' satisfaction and, as a consequence, perceived accuracy of the system.

Knijnenburg et al. (2012) also provide a framework for the user-centered evaluation of recommender systems that extends the system accuracy metric with other relevant measures. For instance, the authors observe correlations between concepts, such as perceived recommendation quality (relevance), choice satisfaction, variety, diversity, effectiveness, and accuracy along with personal characteristics of the user (e.g., trust towards ICT).

To sum up, these types of evaluation criteria focus on subjective user perceptions' in the evaluation of objective aspects of the system. In this article, we do not question the effectiveness of the designed system and its elements, but rather focus on investigating scholars' attitudes towards the recommended people as potential collaborators. The evaluation criteria proposed by the prior research has proven to be effective in the assessment of item recommender systems. In contrast, as objects of recommendation, human individuals contain much more diverse features that influence the evaluation. When assisting people in choosing potential collaborators, the subjective perception of relevance might have various facets and be determined by the need or task for partnering, and, therefore, the metrics should be operationalized accordingly. In this article, we approach relevance criteria with the temporal aspect (i.e., from the perspective of past vs. current research interests), as well as from the perspective of potential collaborative activities with the recommendations.

3. System design

In this section, we first explicate the choice of the data source used for the design of the recommendation algorithm. Next, we outline the data cleaning process and analysis and, finally, describe the user interface developed for the experiment.

3.1. Data source, data cleaning, and analysis

We designed a content-based people recommender system using DBLP, an open bibliographic database of publications records from the majority of Computer Science conferences and journals. The DBLP dataset is a substantial plain ASCII XML file.[4] The metadata for each record contains more than necessary details for the study and, therefore, requires multiple cleaning procedures.

In the first step, the XML file was parsed using the 'xml.sax'[5] package targeting on the following tags: article, inproceedings, proceedings, book, incollection, phdthesis, mastersthesis, and www. Then, from the parsed XML file, we extracted 5,847,090 records that consist only of titles, co-authors, publishing years and venues. Next, in the resulted subset we cleaned the titles of publications following three steps: **(i)** converting letters to lowercase, **(ii)** removing the English stop words with 'nltk.corpus.stopwords'[6] function, **(iii)** removing the digital strings.

The people recommender system runs on the subset of the parsed 5,847,090 records and depends on the input of publication venues of a given user (participant of the study). The detailed data analysis is demonstrated in Figure 1. For a participant, top venues of their publications are given (see step 1 in Figure 1). Then a subset of records is extracted from the parsed DBLP dataset by only those publication venues (step 2). All the titles in the subset records are cleaned as described previously and aggregated to form the corpus profile for each author ever published in those venues (step 3). Those authors who have less than three

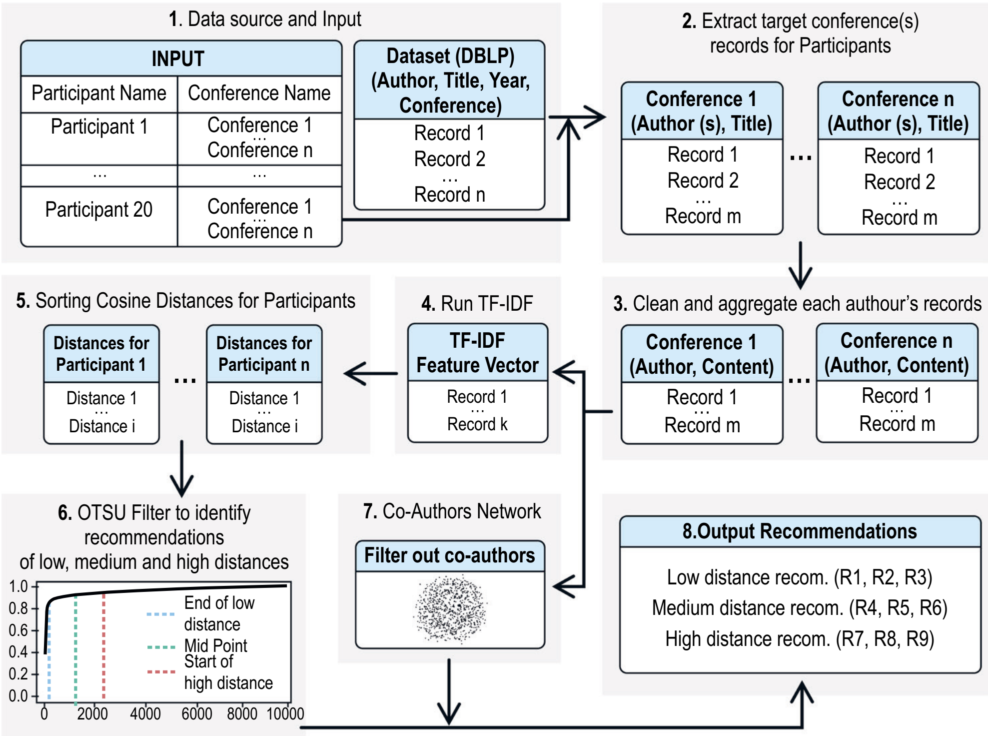


Figure 1. Overview of the DBLP data analysis process to produce people recommendations.

publications in recent five years are filtered out to improve the quality of recommendations. Next, we tokenize the corpus and build the vocabulary with the words that only appear once or show up in more than 95% of the author corpus profiles using 'CountVectorizer()' function from scikit-learn.[7]

After corpus tokenization, TF-IDF is applied to the profile model to form feature vectors for each author (step 4). Next, we compute the cosine distances between the given participant and the other authors in the subset records regarding them (step 5). As it is more intuitive to indicate the close distance with a smaller number, we use cosine distance to represent the similarity between two authors. Accordingly, the closest, or the most similar author to a participant will have the smallest cosine distance.

To validate the participants' preferences on similarity-difference continuum during the user study, we decided to deliver recommendations in the form of three groups of controlled distances – low (high similarity), medium (moderate similar) and high (low similarity). To automatically separate recommendations into such groups, the cosine distances between a participant and the other authors are sorted first, and then the OTSU filter (Otsu 1979) is employed to detect the boundaries between each group (step 6). The OTSU filter calculates the optimum threshold separating the two groups so that their intra-class variance is minimal. As the distribution of the cosine distances follows the power law, we implement the OTSU filter twice. In the first round, we apply it to the whole sorted cosine distances to detect the boundary between the low distance group and the rest. In the second round, we apply the OTSU filter on the rest without the low distance group to divide medium and high distances groups. Finally, three recommendations that have no co-authorship with the given participant (step 7) and published in the same venues are picked from each distance group as the final output, thus delivering nine recommendations in total (step 8).

3.2. User interface

A single-page web application was deployed in Firebase development platform.[8] Figure 2 illustrates the User Interface (UI) view with personalized recommendations in the form of a carousel-based list. The UI visualizes all information about authors, which DBLP data set allows to extract: full name, research topics, the list of co-authors, and recent publications. Each section of the UI is expandable if there is additional content available. The publication list represents only works from conferences where both the recommended person and the participant of the experiment have published in. Accordingly, the list of co-authors is taken from those publications only. The topics were generated through bigram analysis on the corpus profiles of each recommended person. We first generate the bigram word pairs using NLTK 'bigram' function on all the corpus. Next, we use 'nltk.ConditionalFreqDist' to calculate the occurrence of other words by giving a certain word in the corpus. For example, in a bigram word pairs for a word 'social,' the word 'media' may appear 20 times, while the word 'compute' may appear zero times. Then, to generate the authors' topics, the bigram word pairs are

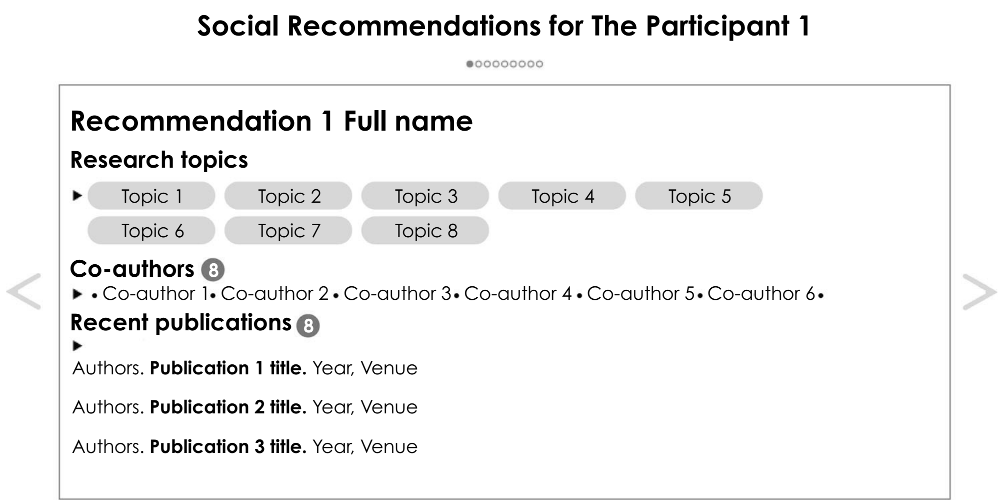


Figure 2. The UI presented to the participant (here with anonymized recommendation).

created from their corpus. After that, we check the conditional frequency of the second word regarding the first word in each pair. If the frequency is equal to or higher than ten (10), we pick this bigram word pair as one of the authors' topics.

4. User study

We designed a user study combining a controlled experiment and a semi-structured interview. By providing participants with real recommendations, we aimed to help them to form their opinion regarding experiment variables. Following the homophily bias, we hypothesize that the lower the cosine distance between the participant and the recommended person (i.e., the similarity of publishing history), the more relevant and similar the recommendation would be perceived.

4.1. Experimental design

In the experiment, the computed cosine distance (content-based distance) is the independent variable, represented as three groups of fellow academics – those with low, medium or high distances. Thus, recommendations of other researchers with *high similarity (low distance)*, *moderate similarity (medium distance)* and *low similarity (high distance)*, with three recommendations from each group were presented to the participants. The participants were not informed of the three groups to avoid biased evaluation, and the presentation order of the altogether nine recommendations was randomized. The evaluation inquired the participants' perceptions about the following dependent variables: relevance, similarity, familiarity and willingness to interact.

4.2. Recruitment and participants

For the experiment, we recruited 18 English-speaking senior researchers who work at two university campuses in Tampere, Finland. Following the assumption that senior researchers often have more needs for finding collaborators, we limited our scope to postdoctoral researchers, professors, or otherwise senior academic positions. For the recruitment, we utilized various e-mail lists to reach relevant faculties, departments, and research groups. In addition to offering the participants a movie ticket for their participation, the recommendations of potential collaborators were also marketed as incentives to take part in the study.

Overall, we had 13 male and five female participants, all based in either of the two universities in the same city. Fourteen of them are Finnish, two Russians, one British, and one Romanian. The ages vary from 32 to 66 (Median: 42, Mean: 45). Seven of the respondents reported their current occupation as Senior Researchers, six as Postdoctoral Researchers, four as Full Professors, and one as an Associate Professor. The most frequent research interest of the participants included human-computer interaction (10), gaze technologies and interactions (7), wireless technologies (6), interaction design and techniques (6), interfaces and information systems (5), usability/user experience and user-centered design (5), telecommunications and networking (5), virtual reality (3), wellness/health technologies (3). Their academic experience varied from 10 to 46 years (Median: 19, Mean: 20.3).

Figure 3 illustrates the participants' backgrounds and attitudes concerning technology orientation, social openness, activity in networking, and breadth of research interests. Along with the other background information, the figure implies that the respondents represent what we would consider as typical computer science scholars, being technically oriented and curious about research, while displaying variety in their networking practices and interests.

4.3. Procedure and data gathering

The data gathering is comprised of three parts: **(i)** screening of suitable participants based on their professional position and publishing history before the experiment session. These data were used to prepare personalized recommendations for each participant. **(ii)** In the experiment session, the participant signed a consent form by filling out an online survey, including also a background questionnaire and numerical evaluations of each recommendation. **(iii)** The experiment was followed by a semi-structured interview to gather qualitative data about the participants' choices and needs for collaboration. The whole study session lasted from 40 minutes to 1.5 hours, depending on the time the participant

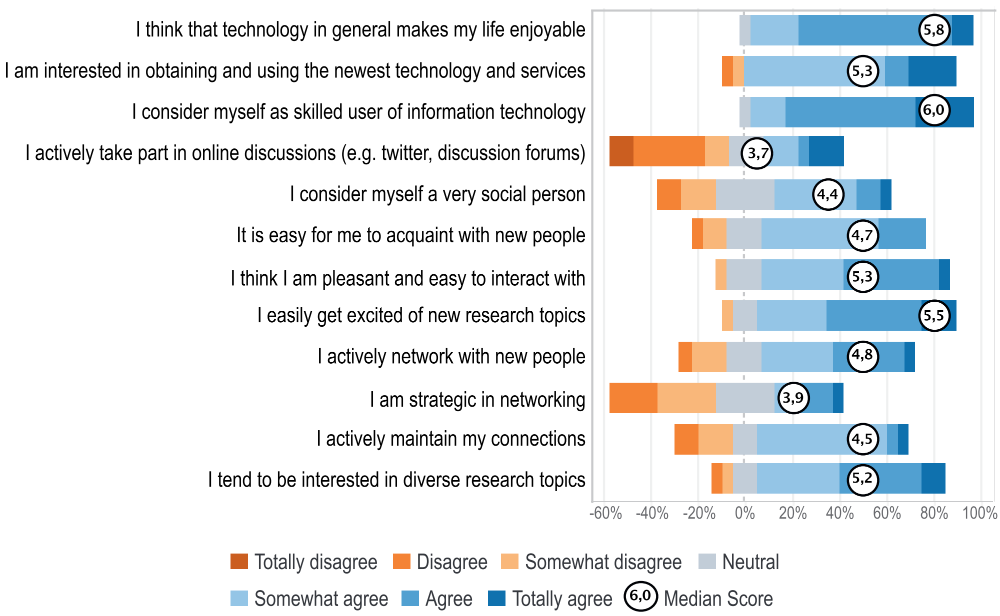


Figure 3. The overview the participants' backgrounds and attitudes.

took to get familiar with and assess the recommendations and how opinionated and expressive they were in the interviewing part. All sessions were audio recorded with the participants' permissions.

Before starting the numerical evaluation, participants were given time to explore all the recommendations and get a general overview of the alternatives. The evaluation was constructed according to four variables (see questionnaire verbatim in Table 1): **(i) perceived relevance** from the perspective of current and past research interests (**Q1** and **Q2**), **(ii)** expected **willingness to interact** with a recommended person in the context of a scientific conference, including six traditional collaborative activities (**Q3-Q8**), **(iii)** levels of **perceived familiarity** – whether or not the user is familiar with the target person, with their research, or with their co-authors (**Q9-Q11**), **(iv) perceived similarity** between the participant and a recommended person (**Q12**). The variables were operationalized based on the authors' personal experiences and qualitative research insights on academic collaboration and user experience evaluation. Originally over 20 candidate items were assessed within the project team and with collaborators in an iterative fashion, resulting in the included 12 items. After providing the ratings, participants were asked to explain the scores and their reasoning behind them verbally. The interview questions that are also presented in Table 1 were designed to obtain participants' rationale behind the scoring of recommendations as well as to reveal needs and factors that affect decision-making in academic networking practices.

4.4. Data analysis

Tableau[9] was used for analysis and visualization of participants' background information, and RStudio[10] for statistical analysis and visualizing the scores in multiple box plots. The experiment has a thrice-repeated within-subjects design with nine categorical data points per participants. We utilized non-parametric Friedman test (Sheldon et al. 1996) and post-hoc analysis with 'Agricolae' package[11] in RStudio to identify a statistically significant difference between the participants' ratings of the three recommendations' groups in all questions. To avoid pseudo-replication, we calculated medians of scores given to each group of recommendations. Thus, the input data for the Friedman test consisted of Participant ID, Similarity distance groups (Low, Medium, High) as factors, and medians of scores as values.

As for the qualitative data, the audio recordings from each session were transcribed and resulted in a text file for each participant (with Min 211, Max 1,373 and Median 690 words). The coding procedure consisted of two cycles including elemental, axial and focused methods (Saldaña 2015). At the first cycle, we applied structural coding that allowed us to group data under top-level categories from interview questionnaire (see Table 1): overall impression, collaboration needs, comments about recommended people and their content, essential factors

Table 1. Online survey structure and questions verbatim.

Recommendations evaluation (Likert 7 point scale, 1: completely disagree, 7: completely agree)

(Perceived Relevance) I consider this person is relevant to me from the perspective of:

Q1. My current research interests

Q2. My past research interests

(Willingness to interact) I consider this person is relevant to me for the following activities at the context of a conference

Q3. Asking for advice

Q4. Giving advice

Q5. Sharing research results and/or ideas

Q6. Exploring joint research interests

Q7. Spending time at a conference together

Q8. Organizing a research visit

Which other activities would this person be interesting to interact with?

(Perceived Familiarity) To what extent are you familiar with this person?

Q9. I am familiar with the research of this person

Q10. I know by name some of the coauthors of this person

Q11. I know him/her in person

(Perceived Similarity) Q12. How similar to you do you find this person's research areas? Very different (1) - Very similar (7)

Interview questions

1. Please tell about your overall impression about the recommended people.
2. Could you please specify what kind of needs for collaboration you normally have?
3. Choose one the most relevant and irrelevant persons and briefly explain why you gave them the scores you did.
4. About whom was it particularly hard to make an evaluation?
5. Which of the presented information about the recommended person do you consider relevant/meaningful in professional matching?
6. What would other factors about the people be important when considering with whom to collaborate?
7. Overall, how would you feel if a system like this gave you recommendations of whom to collaborate with?
8. Would you prefer to receive automatic notifications during a professional event or look for relevant people by yourself?

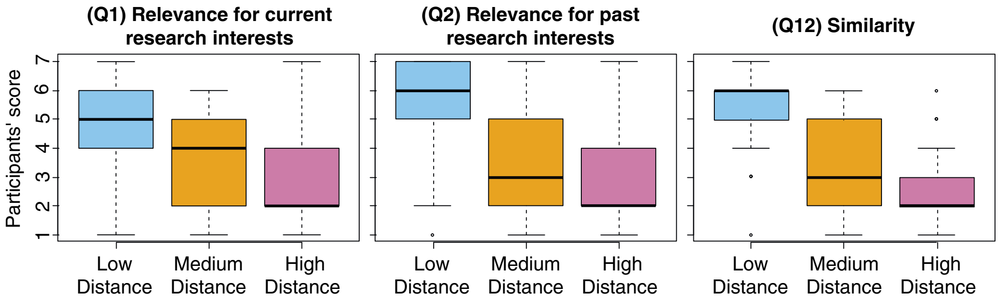


Figure 4. The distribution of participants' scores regarding perceived relevance and similarity.

in social matching, attitudes towards ICT-mediated professional social matching. Then, for the data in each category, we utilized line-by-line analysis and deconstruction of data into emerging categories, which were further reconstituted resulting in subcategories, linkages, and relationships. Finally, the focused coding was applied to identify the most frequent codes and organize them into emerging themes.

5. Results

We first report the quantitative and qualitative results on perceived relevance, similarity, familiarity and willingness to interact. Then, we discuss the needs and important factors in research collaboration, which might have affected the participants' decision-making on potential social interactions.

5.1. Perceived relevance and similarity

5.1.1. Quantitative findings

The participants evaluated the perceived relevance of given recommendations from two perspectives (see Figure 4): relevance for current **(Q1)** and past research interests **(Q2)**. To summarize, the scores were found to be **consistent** with the computer-defined cosine distance. The highest scores were given to the group of recommendations with low distance, while those with the high distance generally received the lowest grades. This indicates the prevalence of homophily bias (preferring most similar researchers) in this sample of participants. Nevertheless, the data also reveals high ratings of perceived relevance for the group of recommendations with low similarity. These matches have received scores of five and higher (appears 16 times) that indicates some participants' interest in dissimilarity and openness towards new opportunities.

A Friedman test and post-hoc analysis indicated a statistically significant difference in the scores of each group of recommendations (see Figure 5). Ratings of relevance for the past and current research interests demonstrated almost equal

Question	Test Statistics	Distance group	Ranks	Friedman's Post Hoc Analysis	
				Low Distance	Medium Distance
Relevance for current research interests (Q1)	N=18, Chi-Square=16.5 DF = 2, p<.001***	Low Distance	47.5		
		Medium Distance	35	<.01**	
		High Distance	25.5	<.001***	<.05*
Relevance fo past research interests (Q2)	N=18, Chi-Square=24.03 DF = 2, p<.001***	Low Distance	50		
		Medium Distance	34.5	<.001***	
		High Distance	23.5	<.001***	<.01**
Similarity (Q12)	N=18, Chi-Square=25,21 DF = 2, p<001***	Low Distance	52		
		Medium Distance	32	<.001***	
		High Distance	24	<.001***	<.05*

Figure 5. Results of Friedman's Test and post-hoc analysis. Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05.

results, meaning that participants were consistent in their scores independently of the temporal perspective.

The quantitative results of perceived similarity **(Q12)** also demonstrate the tendency of a significant difference in given scores (See Figure 4). Accordingly, matches with low distance are graded as most similar, matches with medium distance as somewhat similar and those with high distance – least similar. This consolidates validity of cosine distances and OTSU filter as a method for identifying thresholds of three degrees of similarity. Friedman test results consolidate a significant difference between recommendations groups (See Figure 5).

5.1.2. Qualitative findings

The verbal feedback about relevance and similarity is generally in line with the quantitative results. To provide an overview of participants' comments, we collected illustrative examples in Table 2, sorted according to the three groups of similarity distance. Although the participants were unaware of the three different similarity distances, in their feedback they distinguish between different degrees of perceived relevance by using phrases like *'very/most relevant,' 'somewhat relevant/not exact match'* and *'irrelevant/totally irrelevant.'*

Feedback about the outliers (in Q1 – 9 cases, in Q2 – 7 cases) – recommendations of high distance rated as relevant – has revealed that such recommendations relate to participants' current research interests with potential for future directions, or because of surprising topics appeared in their profiles. Such cases hint about interest in dissimilarity and openness for new opportunities, as quotes 1 and 2 illustrate.

1)"(R8) This person has really interesting research topics. [...] I thought it would be fun to get to know something else, something different from my own research focus. I would definitely go to see their poster. Maybe it would be interesting for my future research. This is an interesting and fun recommendation!" (P1, 50 y.o., Finnish female, Post-doc)

Table 2. Examples of participants' verbal feedback about given recommendations.

Low distance (high similarity)	Medium distance (moderate similarity)	High distance (low similarity)
(R2) I am familiar with this person and his coauthors but not in person because I didn't collaborate with this Greek team directly. However, I was reviewing their publications, particularly, the most recent paper on the list. I consider this recommendation is the most relevant to me. I would surely like to meet in person. (P3, 32 y.o., Russian male, Senior Researcher)	**(R4)** I guess this person is some research leader or professor. Some papers are close to what I am doing, but not an exact match. I would probably find a lot of interesting discussion with him. He seems to be more experienced than me. There is a clear shared interest. I do not remember hearing about him and do not recognize any of co-authors. (P8, 33 y.o., Finnish male, Postdoctoral Researcher)	**(R7)** I have never seen this person before. There are thousands of young researchers, I am not surprised. I have no idea why the system recommends me this person because his topics are totally different from mine. I do not know any of co-authors and topics do not seem to be relevant. (P7, 56 y.o., Finnish male, Senior Researcher)
P3 research interests: wireless technologies, telecommunication and networking	**P8 research interests:** Intenet-of-Things, wireless technology, radio networks, and positioning	**P7 research interests:** Human-computer interaction, VR, gaze technologies and interactions
R2 research topics: wireless technology, cellular systems, sensor networks	**R4 research topics:** Neural networks, stochastic decoding	**R7 research topics:** Crowd science, human-computer interaction

> 2)"(R9) This person is from a different field, and the research has interesting aspects. The last paper in the list is the most interesting: it is about something similar we have been doing recently but not published yet! So, I have to check it and maybe contact the authors." (P2, 42 y.o, Finnish male, Senior Researcher)

Some participants also mentioned that their first impression about recommended people had changed when they started to check the recommendation's profiles in details (see quote 3).

> 3) "That was a good idea that you gave me to check all recommendations first because initial reaction was different, but when I started thinking and realized that my first impression maybe was not correct. When you start thinking about recommendations' relevance further, there might be some changes. So some of them are not that irrelevant as I thought at first." (P16, 50 y.o., Finnish female, Senior Researcher)

Hence, decision-making on perceived relevance was found to be influenced by the participants' estimation about how topics of recommended people match with their own – whether they are similar, very different or complementary. In this regard, few participants addressed that evaluating the similarity between them and the recommended people is a challenging task. First, they mentioned employing different scales of comparison (5 cases) – assessing a recommended person from the perspective of all research fields in the world or specific focus areas. The second factor refers to the temporal aspect (5 cases): as interests and research directions tend to change, the evaluation of the similarity depends on the chosen time frame. Participants also pointed out that the estimation of relevance and optimal level of similarity for specific collaboration are context dependent (4 cases): some tasks would require cooperation with diverse people, while other tasks benefit from similarity. Thus, when looking for candidates with a distinct set of skills, people should also define shared interests or goals to make a prospective collaboration fruitful. Participants also acknowledged the possible adverse effects of receiving recommendations of people who are very similar (4 cases). For example, in a professional context, high similarity of interests might result in competition, and social interaction with such people will require choosing specific communication strategies.

5.2. Familiarity

5.2.1. Quantitative findings

Familiarity variable was evaluated from three perspectives: familiarity with the research topics **(Q9)**; recognizing some of the co-authors' names **(Q10)**; and knowing the match in person **(Q11)**. The Figure 6 depicts that the scores of recommendations with a low distance in all questions are widely distributed, while groups of the medium and high distances received mostly negative ratings. According to scores distribution, participants are mainly familiar with the research of recommendations and aware of many co-authors of recommended people in low distance group. Besides, there are only a few people whom participants know in person. The results of the Friedman test yield a statistically significant difference in given scores (see Figure 7). The post-hoc analysis reveals significant variance only in groups of low vs. medium and low vs. high distances.

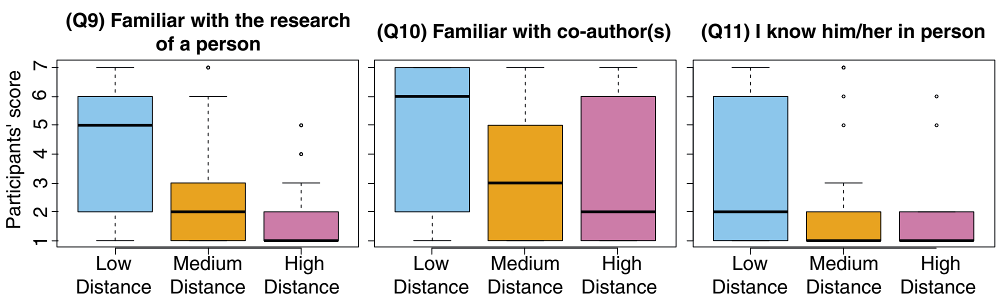


Figure 6. The distribution of participants scores regarding perceived familiarity.

Question	Test Statistics	Distance group	Ranks	Friedman's Post Hoc Analysis: Low Distance	Friedman's Post Hoc Analysis: Medium Distance
Familiar with the research of a person (Q9)	N=18, Chi-Square=21.26 DF = 2, p<.001***	Low Distance	48		
		Medium Distance	32.5	<.001***	
		High Distance	27.5	<.001***	.1108
Familiar with co-author(s) (Q10)	N=18, Chi-Square=12.04 DF = 2, p<.01**	Low Distance	46		
		Medium Distance	33.5	<.01**	
		High Distance	28.5	<.001***	.2597
Know match in person (Q11)	N=18, Chi-Square=23.4 DF = 2, p<.001***	Low Distance	48		
		Medium Distance	30.5	<.001***	
		High Distance	29.5	<.001***	.7048

Figure 7. Results of Friedman's Test and post-hoc analysis. Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05.

5.2.2. *Qualitative findings*

In general, participants were positively surprised to see many unfamiliar people in the list of recommendations and were happy to increase their awareness about new researchers in related scientific fields (see quote 4 and 5). At the same time, they appreciated receiving suggestions about people with established ties or well-known figures and interpreted it as confirmation of the algorithm validity.

> 4) I think the 1st one was most familiar, he is a good match. For the others, I started to wonder about their research topics. I tried to see co-authors identifying relation to me and then to topics. I think it is good that the majority of recommendations are unfamiliar because it is a people recommender system. (P6, 35 y.o., Finnish male, Postdoctoral Researcher)

> 5) I was curious about why I do not know some of the people. I was trying to analyze, and I think there might be a time delay in our academic activities or maybe there is a gap regarding research communities. (P3, 32 y.o., Russian male, Senior Researcher)

In the interview, participants addressed that already known people are less exciting recommendations (quotes 6 and 7). Even though we intentionally filtered out all the co-authors, the bibliographic data prevents understanding of the actual social relationships between researchers. Thus, in some cases (9 recommendations out of 54), the system recommended people from very close social circles, for instance, peers from academic projects or colleagues with whom one had not co-authored publications but interact daily (quote 8).

> 6) "(R3) It seems like I know him. His areas of research match with mine, I would say, by 90%. I know some of his co-authors even in person. This is the most interesting recommendation but not surprising, because I know him and even know his Ph.D. students." (P5, 42 y.o., Finnish male, Full Professor)

> 7) "The first and last person in the list I already know. They are obviously relevant recommendations. But the rest is more interesting because I do not know them." (P10, 40 y.o., Finnish male, Senior Researcher)

8) "(R2) One interesting observation is that one of the recommended people is my roommate. We have similar topics and plenty of other commonalities. We are cooperating mostly by talking. I think, we have never written papers together, but we have shared co-authors." (P4, 56 y.o., Russian male, Senior Researcher).

5.3. Willingness to interact

5.3.1. Quantitative findings

Evaluation of the willingness to interact with recommended people reflects six predefined scenarios of face-to-face interaction or follow-up collaboration at the context of a conference (see Figure 8): **(Q3)** asking advice, **(Q4)** giving advice, **(Q5)** sharing research ideas, **(Q6)** exploring joint research topics, **(Q7)** spending time together, and **(Q8)** organizing a research visit. The distribution of scores demonstrates that participants seemed to have a very positive attitude towards engaging in a low-threshold interaction like sharing research ideas, exploring common research topics and spending time together, particularly with most similar people. Scores given to medium and high distance groups illustrate variance of opinion with a neutral attitude on average. Interestingly, all the six interaction scenarios yielded very similar results, even in the more long-term follow-up action of organizing a research visit. The results of the Friedman test (see Figure 9) depicts a statistically significant difference in given scores only in groups of low vs. high and low vs. medium distance.

5.3.2. Qualitative findings

In general, participants have diversified opinions regarding the estimation of any interactions and follow-up activities. The majority of participants (12 cases)

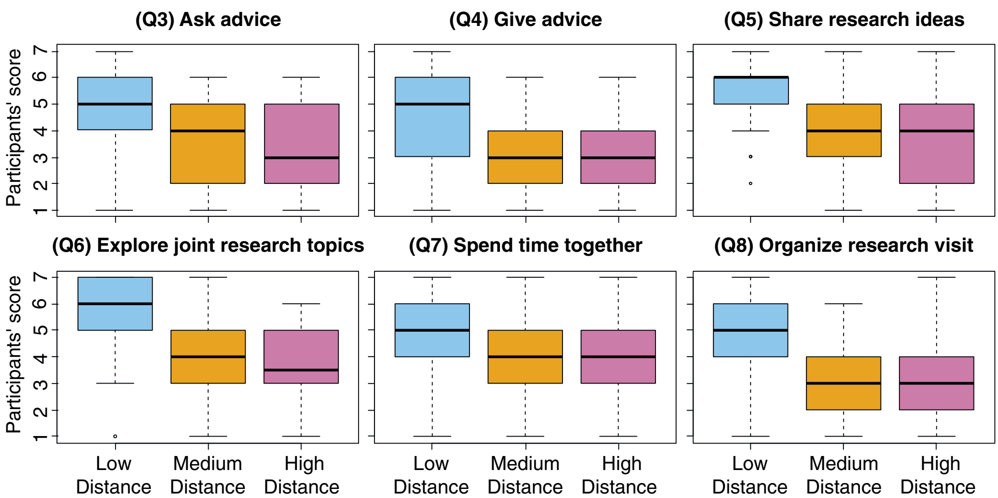


Figure 8. The boxplots visualize participants scores distribution about scenarios of willingness to interact.

Question	Test Statistics	Distance group	Ranks	Friedman's Post Hoc Analysis	
				Low Distance	Medium Distance
Ask advice (Q3)	N=18, Chi-Square=13.54 DF = 2, p<.001***	Low Distance	47		
		Medium Distance	32.5	<.01**	
		High Distance	28.5	<.001***	.3589
Give advice (Q4)	N=18, Chi-Square=9.17 DF = 2, p<.01**	Low Distance	45		
		Medium Distance	31.5	<.01**	
		High Distance	31.5	<.01**	1
Share research ideas (Q5)	N=18, Chi-Square=18.453 DF = 2, p<.001***	Low Distance	48.5		
		Medium Distance	32	<.001***	
		High Distance	27.5	<.001***	.2321
Explore joint research topics (Q6)	N=18, Chi-Square=20.818 DF = 2, p<.001***	Low Distance	50.5		
		Medium Distance	32.5	<.001***	
		High Distance	25	<.001***	.059
Spend time together (Q7)	N=18, Chi-Square=16.754 DF = 2, p<.001***	Low Distance	49		
		Medium Distance	30.5	<.001***	
		High Distance	28.5	<.001***	.633
Organize research visit (Q8)	N=18, Chi-Square=22.286 DF = 2, p<.001***	Low Distance	50		
		Medium Distance	32	<.001***	
		High Distance	26	<.001***	.0831

Figure 9. Results of Friedman's Test and post-hoc analysis. Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05.

consider such social activities to happen naturally at conferences and do not necessarily require high investments of time in collaboration. For some, it was hard to envision willingness to interact with unfamiliar people based on papers titles and topics of the research. Thus, some participants (8 cases) admitted that they would like to learn more about a recommended person before taking any decisions on social interaction. Some emphasize (5 cases) that even in a real context of visiting the conference, it might be challenging to find relevant people in a crowd and contextualized use of such recommender systems might simplify the process and encourage social interaction with unfamiliar people (see quote 9).

> 9) "Such system should narrow the focus to serve specific purposes [...] I think about a scenario where I am going to the conference, so then I can define 'show me people who are relevant to this event' and it would give me a sense of community around it [...] After all these years you sometimes stand somewhere in the corner of a conference hall not knowing anybody. Of course, I can start communicating with random people, but it would be much more efficient if the system can suggest already somewhat relevant people and provide with tickets to talk." (P11, 53 y.o., Finnish female, Full professor)

5.4. Needs and important factors in research collaboration

When specifying crucial needs for collaboration, participants mentioned different activities. The most frequent reasons are understandable when considering senior academics: seeking academic and industrial partners for funding applications (appears in 12 answers) and knowledge sharing as the way of indirect cooperation (12 answers, e.g., exchange of data or finding relevant publications in the topic of interest). Some participants also pointed out a need for people without conflicts of interest (5 answers) such as pre-examiners, reviewers, editors or opponents, who are highly demanded and complicated to find. Another reason was research mobility (5 answers), which calls for cooperation with particularly international universities or companies.

Interestingly, whereas similarity was generally considered to be a significant aspect, the above-mentioned collaborative relationships demand heterogeneity of methodological skills, research areas, or social networks. The participants also emphasized that needs for collaboration are occasional and it will be useful to contextualize the recommender system to specific scenarios, for instance, make it particularly location- and event-based or expert-finder. In their opinion, this will ensure the reasoning for using a service and motivate to follow-up on recommendations (see quote 10).

> 10)"I would be interested in such a system to explore people who are visiting the same conference in advance and filter them based on similarity or relevance. [...] It will help me to revise recommendations faster. Let's say for the event-based mobile application it will be great to inform me when a person visits the event and recommend me to meet him there. If a notification to interact comes in the middle of a street, then I doubt it will work. However, if it will happen at the conference I, of course, will try to follow-up on recommendations." (P8, 33 y.o., Finnish male, Postdoctoral Researcher)

The participants also specified factors that matter to them when seeking professional collaboration. First, the majority (14 replies) addressed the importance of affiliation and the current position of candidates. From their perspective, it can tell a lot about the seniority, availability, and potential interest of the people. Besides, considering the relatively high migration of researchers to non-academic positions, it can indicate whether potential cooperators are still pursuing an academic career. Furthermore, many factors can be implicitly obtained from publications. For instance, the quantity, citation rates, and quality of papers might reveal information about the maturity of a researcher, their topics of interests as well as information about their community. For many (8 replies), these aspects play a significant role when aiming to approach unfamiliar scientists (see quote 11).

> 11) "There are different influence groups with leading experts which are often competitors. So, based on co-authors of a match in his publication lists I can instantly interpret that he belongs to particular influence group. That can help in decision-making whether to collaborate or not and carefully choose the communication strategy." (P5, 42 y.o., Finnish male, Full Professor)

Discussion on seniority level brought out various opinions (7 replies). In general, seniority plays a considerable role: for instance, in tasks that require straightforward ability to make decisions (e.g., project planning) it is essential to be in contact with mature researchers, while some practical implementations could be performed in cooperation with students (e.g., assisting a course). Other scenarios might call for open-mindedness regarding this aspect, like in the following example:

> 12) "The seniority level does not matter to me that much. Sometimes junior people are more creative and innovative. [...] So we should never think about seniority levels. More senior people might have much information, but at the same time too narrow in their vision and interests. Of course, it depends: for consultancy, I might prefer to contact senior people, while for generating new ideas and brainstorming I will be more interested in collaborating with young researchers." *(P15, 50 y.o., Finnish female, Senior Researcher)*

Additionally, participants address the personal chemistry factor (7 replies), specifically for cases of direct collaboration it might be crucial regarding the efficiency of interpersonal relationships and teamwork (see quote 13):

> (13) Chemistry plays a significant role – we need some basis for communication. It should be a person with whom it is nice to sit talk and drink coffee in addition to work practicalities. (P15, 65 y.o., British male, Senior Researcher)

Thus, participants highlighted that this factor is highly demanded yet unfeasible to be integrated into the system because in their opinion personality compatibility can be assessed only after continuous interactions.

6. Discussion

While prior research has aimed at creating meaningful professional connections with the help of people recommender systems, little attention has been put on evaluating the subjective perceptions of the recommendation relevance. We emphasize that recommending is different from predicting new connections (McNee et al. 2006): to design services that can meaningfully enhance professional collaboration, algorithms should go beyond reproducing or strengthening the typical human bias.

In the following, we first summarize our findings and reflect on their novelty and relevance. Next, we provide a discussion on limitations and future work.

6.1. Summary of the results

We presented the results of an experiment on computer science researchers' preferences regarding potential collaborators of different similarity levels with a DBLP-based recommender system. With 18 senior scholars in areas related to CS, we tested how the dependent factors of perceived relevance, similarity, familiarity and willingness to interact are related to the independent variable of objective similarity measurement in terms of publication history.

The findings reveal that (1) the homophily bias is evident also in scholars' intuitive assessments of relevance and willingness to interact, and (2) there is a mismatch between people's intuitive choices and the deliberate intentions in decision-making on potential collaborators.

Considering our first research question about which level of system-defined similarity is preferred in participants' evaluations, the findings demonstrate the highest ratings for most similar people. Methodologically, the subjective evaluations of different similarity levels seem to consolidate the system design, particularly the efficiency of the OTSU filter in identifying different levels on the similarity–difference continuum. In other words, even the relatively simple analytics procedure with scarce data seemed to work sufficiently, and the publication data represented the participants' topics accurately enough. While the norm in such data analytics tends to stress the need for Big Data (e.g., Hoang et al. 2017), it appears that for people recommendations the systems could suffice with rather simple datasets as long as the recommender engine logic is well designed. The findings imply that the participants were able to retrieve useful suggestions and, for the majority, the evaluation process with the operationalized variables was straightforward.

Regarding the second research question about academics' needs and expectations in professional collaboration, the results demonstrate that the optimal area on the similarity-difference continuum highly depends on the type and context of collaboration. For instance, crucial factors in direct cooperation, such as personal compatibility and similarity of attitudes and beliefs, are not as emphasized in short-term and indirect professional interactions (e.g., consultancy type of cooperation) as in long-term collaboration. Furthermore, the nature of the collaboration task might influence the perceived relevance of potential candidates, for example, regarding the complementarity of professional roles, skills, and knowledge.

As a methodological contribution for studying user perceptions, we operationalized the concepts of perceived relevance, similarity, familiarity and willingness to interact as subjective evaluation measures for the context of academic collaboration. This helps to uncover some of the experiential aspects of these concepts and quantitatively assess how individuals consider people recommendations. To complement the reductionist measures, the qualitative findings reveal more complex and nuanced aspects that should be addressed in the design and evaluation of people recommendations for professional partnering.

Following the participants' rationale about important factors in collaboration, we propose that the diversity of recommendations in professional social matching could be enhanced through several dimensions or criteria of relevance:

- *similarity* in terms of background, attitudes, values, beliefs, goals and intentions (e.g., research aims). Previous research addressed that similarity of such qualities can raise cohesion or so-called 'affinity' (Moreland and Zajonc 1982) in interpersonal relationships, and can even reduce adverse effects of individual dissimilarities (Dong et al. 2016) in collaborative work;
- *complementarity* in terms of professional roles, skills, knowledge, and social capital. In this context, complementarity is beyond pure diversity, as discussed by Mitchell and Nicholas (2006). It should enable relevant opportunities for collaboration by identifying beneficial intersections between individuals' qualities;
- *compatibility* for direct cooperation in terms of being mentally, socially, morally or emotionally close to each other. This aspect was partially emphasized by Bozeman et al. (2013), who define collaboration as a process of knowledge production, in which compatibility of such qualities can establish trustful, joyful and personally valued cooperation;
- *approachability/logistics* – the availability of a person for direct or indirect interaction in terms of physical proximity as well as social and organizational distance. This dimension echoes with 'collaboration readiness' conceptualized by Olson and Olson (2000), who calls for better technological solutions to enable smooth communication and interaction practices for distributed collaboration.

6.2. Limitations and future work

We selected a mixed-method approach to enable a broad understanding of our research questions. By combining a controlled experiment with qualitative face-to-face interviews, we intentionally limited the sample size and compromised generalizability with deeper qualitative understanding. In the same vein, the participants represent culturally the same geographical area, which means that the generalizability of the findings to the general population of scholars is limited.

Nevertheless, our method allowed us to observe the decision-making process on collaboration in the actual context of using a people recommender system. Additionally, it helped to engage participants in the discussion on potentially relevant partners concerning similarity, complementarity, and other essential qualities or factors when seeking collaboration. We could not have elicited some of the interview findings without having the task of evaluating the recommendations in situ. In fact, our prior research experience suggests that qualitative exploration of human needs, wishes, and expectations often benefits from providing a design artifact that can help a participant to form opinions on abstract concepts and speculated behavior.

6.2.1. Data set limitations

The limitations on data set might explain some of the mismatches in participants' quantitative and qualitative feedback on recommendations. A central limitation of DBLP is that only the publication titles (without abstracts) are available, which limits the content analysis and compromises the accuracy of the user profiles. For example, some participants addressed that it was problematic to assess the relevance of junior researchers who had a small number of papers (Min 3 in the reported study). Limiting the data sample to only those who have extensive publication history would be advantageous for the accuracy of topic modeling and providing comprehensive pictures of the assessed individuals. However, we intentionally wanted to introduce both junior and senior researchers and reveal how they could be appreciated in different contexts of cooperation, as well as evaluate the role of seniority as a possibly influential factor in collaboration processes.

Another limitation of the DBLP data set is that it provides limited understanding to the social ties between researchers. Even though we excluded all the co-authors of each participant from the recommendations, some were recommended people with whom they occasionally interact (9 cases out of total 54 recommendations). The proportion of very familiar people was relatively small, so this can be argued to have an relatively small effect on the perceptions of the system validity. By applying alternative data sources, if practically feasible, it would be possible to implement more advanced analysis of social networks and, thus, prevent recommending already known people.

6.2.2. Homophily bias

Even though the participants of the experiment were unaware of the similarity distance groups, the quantitative ratings of recommendations indeed demonstrated the tendency of researchers preferring most similar people. Only a few participants assessed recommendations with high distances as exciting, surprising and worthy of exploration for potential follow-up. At the same time, the qualitative feedback provides evidence of aspects in a research collaboration that require access to both similar and different others. In the following, we discuss two possible reasons behind the apparent homophily bias, also related to the methodological validity of the study:

(i) *The evaluation was largely based on first impression.* Studies on the cognitive processes of choice (Kahneman 2003; Stanovich and West 2000) distinguish between two modes influencing humans' decision-making – so-called 'system 1' (effortless, intuition-based judgment) and 'system 2' (rational and reasoning-based). In our experimental setup, it seems that most of the participants were primarily relying on their intuition and did not engage in more rational or reflective reasoning in their evaluation. Therefore, the first impression about recommendations was likely driven by the homophily bias, thus explaining the numerical evaluations. At the same time, after rationalizing the matter during the interview, and reflecting with specific practical collaboration scenarios, the participants started appreciating different types of diversity between themselves and the evaluated person. As for considerations for design, this finding calls for user interfaces that support reflection and multi-dimensional analysis of the collaboration potential with a given recommendation. The current norm in recommender systems and leisurely social matching is based on *hastiness*: using simplistic profiles and simple mechanisms for selecting or discarding recommendations. We argue that such UI and interaction mechanisms do not fit with the goal of identifying optimal academic collaborators.

(ii) *Lacking a timely need for collaboration.* At the moment of the experiment, all of the participants were involved in research projects where the consortium was already built, and they did not report having any urgent needs or requirements for finding new collaborators. Therefore, the estimation of relevance was mostly formed according to their general picture of an ideal collaborator. Nevertheless, the majority emphasized the occasional need to utilize such people recommender systems for research networking and collaboration. This raises two design considerations. First, this calls for user interfaces that support keeping track of different types of collaboration needs in the often so scattered work of academics; most simply, the user could be reminded about their different professional activities upon receiving a recommendation. Second, this calls for context awareness (Mayer et al. 2015b) in timing the recommendation. For example, rather than having separate services for professional matching, the recommendations could be tied to services that academics typically use to seek for suitable collaborators.

7. Conclusions

We evaluated scholars' perceptions of relevance about potential collaborators representing different levels of similarity, utilizing a bibliography-based people recommender system. We operationalized the concept of perceived relevance, familiarity, similarity and willingness to interact within the context of evaluating prospective collaboration. By showing how these variables match with system-defined similarity in bibliography data, we revealed the asymmetry of scholars' intuition-based evaluation and their intentions. The quantitative results demonstrated the effects of homophily bias (preference of most similar others) to perceived relevance, while qualitative findings identify important factors for collaboration that naturally require connection with people of complementary expertise. Compared to the evaluation methods used in item recommenders, the findings demonstrate that people recommender systems require more advanced models and logics that go beyond predicting ties or optimizing for accuracy. From cognitive psychology perspective, assessment of potential partnering with recommended people is a complicated task that should not rely only on the first impression. Considering the long-term and reciprocal nature of professional collaboration, social matching of scholars calls for domain-specific solutions.

Acknowledgments

This work was conducted under Business Finland project Big Match (3166/31/2017 and 3074/31/2017). We owe special gratitude to Tuomas Kaittola for the implementation of the system. We thank all the members of the project who contributed to the design of the experiment.

Notes

1 DBLP database – http://dblp.org/statistics/recordsindblp.html
2 TF-IDF is a numerical statistic that is intended to reflect the importance of a word to a document in a collection or corpus.
3 Cosine distance is a measure of angular similarity between two non-zero vectors of an inner product space that measures the cosine of the angle between them.
4 The DBLP dataset (2.0 G) can be downloaded from the official website – https://dblp.uni-trier.de/xml/
5 XML.SAX package provides a number of modules which implement the Simple API for XML (SAX) interface for Python – https://docs.python.org/2/library/xml.sax.html
6 The Natural Language Toolkit (NLTK) is a Python-based suite of libraries and programs for symbolic and statistical natural language processing. Official NLTK website – http://www.nltk.org
7 Scikit-learn is a tool for data mining and data analysis in Python – http://scikitlearn.org/stable/about.html
8 Firebase is a mobile and web application development platform which operates on Google infrastructure – https://firebase.google.com/
9 Tableau – a data analytics and visualization tool – https://www.tableau.com
10 An open-source integrated development environment for R, a programming language for statistical computing and graphics – https://www.rstudio.com
11 Agricolae documentation – https://www.rdocumentation.org/packages/agricolae/versions/1.2-8

References

Ackerman, Mark S.; and David W. McDonald (1996). Answer Garden 2: merging organizational memory with collaborative help. *CSCW'96.Proceedings of the ACM Conference on Computer Supported Cooperative Work, Boston, Massachusetts, USA, 16–20 November 1996*. New York: ACM, pp. 97–105.

Adamopoulos, Panagiotis; and Alexander Tuzhilin (2015). On unexpectedness in recommender systems: Or how to better expect the unexpected. *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 5, no. 4, January 2015, pp. 54.

Arens-Volland, Andreas; and Yannick Naudet (2016). Personalized recommender system for event attendees. *SMAP'16. International Workshop on Semantic and Social Media Adaptation and Personalization, Thessaloniki, Greece, 20-21 October 2016*. IEEE, pp. 65–70.

Argote, Linda; and Ron Ophir (2002). Intraorganizational learning. In: J. A. C. Baum (ed.): *The Blackwell Companion to Organizations*. Hoboken, New Jersey: Blackwell Publishers Ltd., Chapt. 8, pp. 181–207.

Beel, Joeran; Bela Gipp; Stefan Langer; and Corinna Breitinger (2016). Research-paper recommender systems: a literature survey. *International Journal on Digital Libraries*, vol. 17, no. 4, November 2016, pp. 305–338.

Beham, Günter; Barbara Kump; Tobias Ley; and Stefanie Lindstaedt (2010). Recommending knowledgeable people in a work-integrated learning system. *Procedia Computer Science*, vol. 1, no. 2, September 2010, pp. 2783–2792.

Bessi, Alessandro (2016). Personality traits and echo chambers on facebook. *Computers in Human Behavior*, vol. 65, December 2016, pp. 319–324.

Bozeman, Barry; and Elizabeth Corley (2004). Scientists' collaboration strategies: implications for scientific and technical human capital. *Research Policy*, vol. 33, no. 4, April 2004, pp. 599–616.

Bozeman, Barry; Daniel Fay; and Catherine P. Slade (2013). Research collaboration in universities and academic entrepreneurship: the-state-of-the-art. *The Journal of Technology Transfer*, vol. 38, no. 1, February 2013, pp. 1–67.

Brusilovsky, Peter; Jung Sun Oh; Claudia López; Denis Parra; and Wei Jeng (2017). Linking information and people in a social system for academic conferences. *New Review of Hypermedia and Multimedia*, vol. 23, no. 2, June 2017, pp. 81–111.

Burt, Ronald S. (2017). Structural holes versus network closure as social capital. In: N. Lin, K. Cook, and R. S. Burt (eds.): *Social capital*. New York: Routledge, Chapt. 2, pp. 31–56.

Börner, Katy; Noshir Contractor; Holly J Falk-Krzesinski; Stephen M. Fiore; Kara L. Hall; Joann Keyton; Bonnie Spring; Daniel Stokols; William Trochim; and Brian Uzzi (2010). A multi-level systems perspective for the science of team science. *Science Translational Medicine*, vol. 2, no. 49, September 2010, pp. 1–5.

Castells, Pablo; Neil J. Hurley; and Saul Vargas (2015). Novelty and diversity in recommender systems. In: F. Ricci, L. Rokach, and B. Shapira (eds.): *Recommender Systems Handbook*. Boston, MA: Springer, Chapt. 26, pp. 881–918.

Chang, Shuo; Vikas Kumar; Eric Gilbert; and Loren G. Terveen (2014). Specialization, homophily, and gender in a social curation site: findings from pinterest. *CSCW'14. Proceedings of the ACM conference on Computer Supported Cooperative Work and Social Computing, Baltimore, Maryland, USA, 15–19 February 2014*. New York: ACM, pp. 674–686.

Chin, Alvin; Bin Xu; Chen Zhao; and Xia Wang (2014). From offline to online: connecting people with a mobile social networking application at a conference. *Chinese CHI '14. Proceedings of the Second International Symposium of Chinese CHI, Toronto, Ontario, Canada, 26–27 April 2014*. New York: ACM, pp. 40–49.

Cox, Donna; Volodymyr Kindratenko; and David Pointer (2003). IntelliBadge TM: towards providing location-aware value-added services at academic conferences. In: A. K. Dey, A. Schmidt, and J. F. McCarthy (eds.): *UbiComp 2003. International Conference on Ubiquitous Computing, Seattle, WA, USA, 12–15 October 2003*, Vol. 2864. Berlin, Heidelberg: Springer, pp. 264–280.

Dey, Anind K.; Daniel Salber; Gregory D. Abowd; and Masayasu Futakawa (1999). The conference assistant: Combining context-awareness with wearable computing. *Digest of Papers. Third International Symposium on Wearable Computers, San Francisco, California, USA, 18–19 October 1999*. IEEE, pp. 21–28.

Dong, Wei; Kate Ehrlich; Michael M. Macy; and Michael Muller (2016). Embracing cultural diversity: Online social ties in distributed workgroups. *CSCW'16. Proceedings of the ACM Conference on Computer Supported Cooperative Work and Social Computing, San Francisco, California, USA, 27 February – 2 March 2016*. New York: ACM, pp. 274–287.

Farzan, Rosta; and Peter Brusilovsky (2007). Community-based conference navigator. *UM 2007. Proceedings of Workshop on Adaptation and Personalisation in Social Systems: Groups, Teams, Communities at the International Conference on User Modeling, Corfu, Greece, 25–29 July 2007*, Vol. 2007. Dordrecht, The Netherlands: Kluwer Academic Publishers, pp. 30–39.

Frydlinger, David; Jeanette Nyden; and Kate Vitasek (2013). Unpacking Collaboration Theory: What Every Negotiator Should Know to Establish Successful Strategic Relationships. http://www.vestedway.com/vested-library. Accessed 3 March 2019.

Graves, Laura M.; and Priscilla M. Elsass (2005). Sex and sex dissimilarity effects in ongoing teams: Some surprising findings. *Human Relations*, vol. 58, no. 2, February 2005, pp. 191–221.

Guy, Ido; and Luiz Pizzato (2016). People Recommendation Tutorial. *RecSys'16. Proceedings of the ACM Conference on Recommender Systems, Boston, Massachusetts, USA, 15–19 September 2016*. New York: ACM, pp. 431–432.

Guy, Ido; Michal Jacovi; Adam Perer; Inbal Ronen; and Erel Uziel (2010). Same places, same things, same people?: mining user similarity on social media. *CSCW'10. Proceedings of the 2010 ACM conference on Computer Supported Cooperative Work, Savannah, Georgia, USA, 06–10 February 2010*. New York: ACM, pp. 41–50.

Heck, Tamara (2013). Combining social information for academic networking. *CSCW '13. Proceedings of the ACM Conference on Computer Supported Cooperative Work, San Antonio, TX, USA, 23–27 February 2013*. New York: ACM, pp. 1387–1398.

Hoang, Dinh Tuyen; Van Cuong Tran; Van Du Nguyen; Ngoc Thanh Nguyen; and Dosam Hwang (2017). Improving Academic Event Recommendation Using Research Similarity and Interaction Strength Between Authors. *Cybernetics and Systems*, vol. 48, no. 3, March 2017, pp. 210–230.

Hobman, Elizabeth V.; Prashant Bordia; and Cynthia Gallois (2004). Perceived dissimilarity and work group involvement: The moderating effects of group openness to diversity. *Group & Organization Management*, vol. 29, no. 5, October 2004, pp. 560–587.

Hsiehchen, David; Magdalena Espinoza; and Antony Hsieh (2015). Multinational teams and diseconomies of scale in collaborative research. *Science advances*, vol. 1, no. 8: e1500211, September 2015, pp. 1–9.

Jasny, Lorien; Joseph Waggle; and Dana R. Fisher (2015). An empirical examination of echo chambers in US climate policy networks. *Nature Climate Change*, vol. 5, no. 8, August 2015, pp. 782–786.

Kahneman, Daniel (2003). A perspective on judgment and choice: mapping bounded rationality. *American psychologist*, vol. 58, no. 9, September 2003, pp. 697.

Kawakita, Ysuke; Shirou Wakayama; Hisakazu Hada; Osamu Nakamura; and Jun Murai (2004). Rendezvous enhancement for conference support system based on RFID. *SAINT 2004. International Symposium on Applications and the Internet Workshops, Tokyo, Japan, 26–30 January 2004*. IEEE, pp. 280–286.

Knijnenburg, Bart P.; Martijn C. Willemsen; Zeno Gantner; Hakan Soncu; and Chris Newell (2012). Explaining the user experience of recommender systems. *User Modeling and User-Adapted Interaction*, vol. 22, no. 4-5, October 2012, pp. 441–504.

Kong, Xiangjie; Huizhen Jiang; Zhuo Yang; Zhenzhen Xu; Feng Xia; and Amr Tolba (2016). Exploiting publication contents and collaboration networks for collaborator recommendation. *PloS one*, vol. 11, no. 2: e0148492, February 2016, pp. 1–13.

Lazarsfeld, Paul F; and Robert K. Merton (1954). Friendship as a social process: A substantive and methodological analysis. *Freedom and control in modern society*, vol. 18, no. 1, pp. 18–66.

Lee, Jae Kook; Jihyang Choi; Cheonsoo Kim; and Yonghwan Kim (2014). Social media, network heterogeneity, and opinion polarization. *Journal of communication*, vol. 64, no. 4, January 2014, pp. 702–722.

Li, Baoli; and Liping Han (2013). Distance weighted cosine similarity measure for text classification. In: H. Yin, K. Tang, Y. Gao, F. Klawonn, M. Lee, T. Weise, B. Li, and X. Yao (eds.): *IDEAL 2013. International Conference on Intelligent Data Engineering and Automated Learning. Lecture Notes in Computer Science, Hefei, China, 20–23 October 2013*. Berlin, Heidelberg: Springer, pp. 611–618.

Li, Lin; Anna Scaglione; Ananthram Swami; and Qing Zhao (2013). Consensus, polarization and clustering of opinions in social networks. *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 6, June 2013, pp. 1072–1083.

Li, Jing; Feng Xia; Wei Wang; Zhen Chen; Nana Yaw Asabere; and Huizhen Jiang (2014). Acrec: a co-authorship based random walk model for academic collaboration recommendation. *WWW'14. Proceedings of the 23rd International Conference on World Wide Web, Seoul, Korea, 07–11 April 2014*. New York: ACM, pp. 1209–1214.

Mannix, Elizabeth; and Margaret A. Neale (2005). What differences make a difference? The promise and reality of diverse teams in organizations. *Psychological Science in the Public Interest*, vol. 6, no. 2, October 2005, pp. 31–55.

Marsden, Peter V. (1987). Core discussion networks of Americans. *American sociological review*, vol. 52, no. 1, February 1987, pp. 122–131.

Mayer, Julia M.; Quentin Jones; and Starr Roxanne Hiltz (2015a). Identifying opportunities for valuable encounters: Toward context-aware social matching systems. *ACM Transactions on Information Systems (TOIS)*, vol. 34, no. 1, October 2015, pp. 1–32.

Mayer, Julia M.; Starr Roxanne Hiltz; and Quentin Jones (2015b). Making Social Matching Context-Aware: Design Concepts and Open Challenges. *CHI'15. Proceedings of the ACM Conference on Human Factors in Computing Systems, Seoul, Republic of Korea, 18–23 April 2015*. New York: ACM, pp. 545–554.

Mayer, Julia M.; Starr Roxanne Hiltz; Louise Barkhuus; Kaisa Väänänen; and Quentin Jones (2016). Supporting opportunities for context-aware social matching: An experience sampling study. *CHI'16. Proceedings of the ACM Conference on Human Factors in Computing Systems, San Jose, California, USA, 07–12 May 2016*. New York: ACM, pp. 2430–2441.

McNee, Sean M.; John Riedl; and Joseph A. Konstan (2006). Being accurate is not enough. *CHI EA'06. Extended abstracts on Human factors in computing systems, Montréal, Québec, Canada, 22–27 April 2006*. New York: ACM Press, pp. 1097–1101.

Mitchell, Rebecca; and Stephen Nicholas (2006). Knowledge creation in groups: The value of cognitive diversity, transactive memory and open-mindedness norms. *The Electronic Journal of Knowledge Management (EJKM)*, vol. 4, no. 1, December 2006, pp. 67–74.

Mollica, Kelly A.; Barbara Gray; and Linda K. Treviño (2003). Racial homophily and its persistence in newcomers' social networks. *Organization Science*, vol. 14, no. 2, March – April 2003, pp. 123–136.

Moody, James (2001). Race, school integration, and friendship segregation in America. *American journal of Sociology*, vol. 107, no. 3, November 2001, pp. 679–716.

Moreland, Richard L.; and Robert B. Zajonc (1982). Exposure effects in person perception: Familiarity, similarity, and attraction. vol. 18, no. 5, September 1982, pp. 395–415.

Muller, Michael; Kate Ehrlich; Tara Matthews; Adam Perer; Inbal Ronen; and Ido Guy (2012). Diversity among enterprise online communities: collaborating, teaming, and innovating through social media. *CHI'12. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Austin, Texas, USA 5–10 May 2012*. New York: ACM, pp. 2815–2824.

Nguyen, Tien T.; Pik-Mai Hui; F. Maxwell Harper; Loren Terveen; and Joseph A. Konstan (2014). Exploring the filter bubble: the effect of using recommender systems on content diversity. *WWW'14. Proceedings of the 23rd international conference on World wide web, Seoul, Korea, 7–11 April 2014*. New York: ACM, pp. 677–686.

Nishibe, Yoshiyasu; Hiroaki Waki; Ichiro Morihara; Fumio Hattori; Toru Ishida; Toshikazu Nishimura; Hirofumi Yamaki; Takaaki Komura; Nobuyasu Itoh; Tadahiro Gotoh; Toyoaki Nishida; Hideaki Takeda; Atsushi Sawada; Harumi Maeda; Masao Kajihara; and Hidekazu Adachi (1998). Mobile digital assistants for community support. *AI Magazine*, vol. 19, no. 2, June 1998, pp. 31.

Olson, Gary M.; and Judith S. Olson (2000). Distance matters. *Human–computer Interaction*, vol. 15, no. 2-3, December 2000, pp. 139–178.

Otsu, Nobuyuki (1979). A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, vol. 9, no. 1, January 1979, pp. 62–66.

Paasovaara, Susanna; Ekaterina Olshannikova; Pradthana Jarusriboonchai; Aris Malapaschas; and Thomas Olsson (2016). Next2You: a proximity-based social application aiming to encourage interaction between nearby people. *MUM'16. Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia, Rovaniemi, Finland, 12–15 December 2016*. New York: ACM, pp. 81–90.

Parra, Denis; Wei Jeng; Peter Brusilovsky; Claudia López; and Shaghayegh Sahebi (2012). Conference Navigator 3: An online social conference support system. *UMAP 2012. UMAP workshops at the conference on User Modeling, Adaptation, and Personalization, Montréal, Québec, 16–20 July 2012*. Dordrecht, The Netherlands: Kluwer Academic Publishers, pp. 1–4.

Pu, Pearl; Li Chen; and Rong Hu (2011). A user-centric evaluation framework for recommender systems. *RecSys'11. Proceedings of the fifth ACM conference on Recommender systems, Chicago, Illinois, USA, 23–27 October 2011*. New York: ACM, pp. 157–164.

Pu, Pearl; Li Chen; and Rong Hu (2012). Evaluating recommender systems from the user's perspective: survey of the state of the art. *User Modeling and User-Adapted Interaction*, vol. 22, no. 4-5, October 2012, pp. 317–355.

Rajagopal, Kamakshi; Jan M. van Bruggen; and Peter B. Sloep (2017). Recommending peers for learning: Matching on dissimilarity in interpretations to provoke breakdown. *British Journal of Educational Technology*, vol. 48, no. 2, February 2017, pp. 385–406.

Rodan, Simon; and Charles Galunic (2004). More than network structure: How knowledge heterogeneity influences managerial performance and innovativeness. *Strategic Management Journal*, vol. 25, no. 6, April 2004, pp. 541–562.

Saldaña, Johnny (2015). *The coding manual for qualitative researchers*. London: SAGE.

Sharma, Amit; and Dan Cosley (2016). Distinguishing between personal preferences and social influence in online activity feeds. *CSCW'16. Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work and Social Computing, San Francisco, California, USA, 27 February – 02 March 2016*. New York: ACM, pp. 1091–1103.

Sheldon, Michael R.; Michael J. Fillyaw; and W. Douglas Thompson (1996). The use and interpretation of the Friedman test in the analysis of ordinal-scale data in repeated measures designs. *Physiotherapy Research International*, vol. 1, no. 4, November 1996, pp. 221–228.

Sie, Rory L. L.; Hendrik Drachsler; Marlies Bitter-Rijpkema; and Peter Sloep (2012). To whom and why should I connect? Co-author recommendation based on powerful and similar peers.

International Journal of Technology Enhanced Learning (IJTEL), vol. 4, no. 1-2, August 2012, pp. 121–137.

Sinha, Rashmi; and Kirsten Swearingen (2002). The role of transparency in recommender systems. *CHI EA'02. CHI'02 Extended Abstracts on Human Factors in Computing Systems, Minneapolis, Minnesota, USA, 20 - 25 April 2002*. New York: ACM, pp. 830–831.

Stanovich, Keith E.; and Richard F. West (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, vol. 23, no. 5, October 2000, pp. 645–665.

Terveen, Loren; and David W. McDonald (2005). Social matching: A framework and research agenda. *ACM transactions on computer-human interaction (TOCHI)*, vol. 12, no. 3, September 2005, pp. 401–434.

Tsai, Chun-Hua; and Peter Brusilovsky (2016). A personalized people recommender system using global search approach. *IConference 2016. Proceedings of IConference, Philadelphia, Pennsylvania, USA, 20–23 March 2016*. iSchools, pp. 1–5.

Vassileva, Julita; Gordon McCalla; and Jim Greer (2003). Multi-agent multi-user modeling in I-Help. *User Modeling and User-Adapted Interaction*, vol. 13, no. 1-2, February 2003, pp. 179–210.

Wongchokprasitti, Chirayu; Peter Brusilovsky; and Denis Parra-Santander (2010). Conference Navigator 2.0: community-based recommendation for academic conferences. *SRS '10. Workshop on Social Reminder Systems, Hong Kong, China 07–07 February 2010*. New York: ACM, pp. 1–5.

Wuchty, Stefan; Benjamin F. Jones; and Brian Uzzi (2007). The increasing dominance of teams in production of knowledge. *Science*, vol. 316, no. 5827, May 2007, pp. 1036–1039.

Ye, Teng; and Lionel P. Robert Jr. (2017). Does collectivism inhibit individual creativity?: The effects of collectivism and perceived diversity on individual creativity and satisfaction in virtual ideation teams. *CSCW'17. Proceedings of the ACM Conference on Computer Supported Cooperative Work and Social Computing, Portland, OR, USA, 25 February – 01 March 2017*. New York: ACM, pp. 2344–2358.

Yuan, Y. Connie; and Geri Gay (2006). Homophily of network ties and bonding and bridging social capital in computer-mediated distributed teams. *Journal of Computer-Mediated Communication*, vol. 11, no. 4, October 2006, pp. 1062–1084.

Zaiane, Osmar R.; Jiyang Chen; and Randy Goebel (2007). DBconnect: mining research community on DBLP data. *WebKDD/SNA-KDD'07. Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis, San Jose, California, 12 August 2007*. New York: ACM, pp. 74–81.

Zenk, Lukas; Michael Smuc; and Florian Windhager (2014). Beyond the name tag. In: B. Lutz (ed.): *Wissen nimmt Gestalt an: Beiträge zu den Kremser Wissensmanagement-Tagen 2013*. Danube, Austria: Edition-Donau-Univ. Krems, pp. 215–224.

Zhang, Amy X.; Anant Bhardwaj; and David Karger (2016). Confer: A Conference Recommendation and Meetup Tool. *CSCW'16 Companion. Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion, San Francisco, California, USA, 26 February – 2 March 2016*. New York: ACM, pp. 118–121.

Publisher's Note

PUBLICATION
V

Utilizing structural network positions to diversify people recommendations on twitter

Olshannikova, E., Skenderi, E., Olsson, T., Koivunen, S., and Huhtamäki, J.,

Advances in Human-Computer Interaction, *2022*(6584394), 17
https://doi.org/10.1155/2022/6584394

Hindawi
Advances in Human-Computer Interaction
Volume 2022, Article ID 6584394, 17 pages
https://doi.org/10.1155/2022/6584394

Research Article

Utilizing Structural Network Positions to Diversify People Recommendations on Twitter

Ekaterina Olshannikova ,[1] Erjon Skenderi ,[2] Thomas Olsson ,[1] Sami Koivunen ,[1] and Jukka Huhtamäki [2]

[1]*Faculty of Information Technology and Communication Studies, Tampere University, Kalevantie 4, Tampere 33100, Finland*
[2]*Department of Industrial and Information Management, Tampere University, Kalevantie 4, Tampere 33100, Finland*

Correspondence should be addressed to Ekaterina Olshannikova; ekaterina.olshannikova@tuni.fi

Received 9 July 2021; Accepted 31 December 2021; Published 30 January 2022

Academic Editor: Christos Troussas

Social recommender systems, such as “Who to follow” on Twitter, utilize approaches that recommend friends of a friend or interest-wise similar people. Such algorithmic approaches have been criticized for resulting in filter bubbles and echo chambers, calling for diversity-enhancing recommendation strategies. Consequently, this article proposes a social diversification strategy for recommending potentially relevant people based on three structural positions in egocentric networks: dormant ties, mentions of mentions, and community membership. In addition to describing our analytical approach, we report an experiment with 39 Twitter users who evaluated 72 recommendations from each proposed network structural position altogether. The users were able to identify relevant connections from all recommendation groups. Yet, perceived familiarity had a strong effect on perceptions of relevance and willingness to follow-up on the recommendations. The proposed strategy contributes to the design of a people recommender system, which exposes users to diverse recommendations and facilitates new social ties in online social networks. In addition, we advance user-centered evaluation methods by proposing measures for subjective perceptions of people recommendations.

1. Introduction

Social media and social networking services such as Twitter are widely used in professional cooperation within and across organizations, helping to gain new insights and share knowledge. The functionality of recommending new connections is essential for expanding the social network and introducing new professional ties. Such *people recommender systems* represent the areas of social computing and social matching [1], which are argued to require careful design of the algorithmic principles [2]. Thus, people recommenders aim at influencing followership by suggesting seemingly suitable others based on user modeling and predictive analytics.

The majority of existing approaches tend to support *homophily* bias [3]—a tendency of preferring others with similar characteristics as oneself, focusing on similarities in user-created content [4]. Another commonly used principle is the *triadic closure* [5] in the followership networks [6] that focuses on friend-of-a-friend connections. Furthermore, the “Who to follow” feature on Twitter has been found to favor already popular users and promote uni-directional network connections [7]. A recently much-discussed concern is that network-based algorithms on social media can lead to echo chambers and perpetuate social polarization [8] because they are efficient in reproducing existing connections but limited in developing new ones. Therefore, introducing new social ties is likely to be based on similarity or close social vicinity of the active user.

Consequently, an important goal has been set to increase diversity in the recommendations [9, 10], potentially decreasing human and algorithmic biases [11]. Our work highlights this goal toward diversification and heterogeneity, especially in the professional networking context where diversity is seen as a key driver for fruitful collaboration [2]. Diversifying people recommendations can enable unexpected yet valuable social encounters [12], which require alternative recommendation strategies to identify relevant people in the vast and complex Twitter network. Traditional recommender systems research seeks to optimize algorithmic accuracy and effectiveness [13, 14], creating algorithms that can reproduce actors' current behavior as accurately as possible [15] rather than aiming at increasing diversity. In turn, focusing on accuracy results in a lack of user-centered research addressing the intricacies of recommendation strategies regarding the desirable degree and types of diversity exposure. To this end, understanding the users' subjective perceptions of the relevance of given diversity-oriented people recommendations is crucial.

An ongoing merger of three nearby universities provided an opportune case study for exploring a new social matching strategy on Twitter. This merger raised a need to enable cross-sectoral collaboration between scholars and stakeholders within the new university community [16]. Prior research suggests that bridging polarized intellectual communities and increasing social awareness contributes to developing creativity and innovation capabilities [17]. The pool of Twitter users following one of the to-be-merged universities represents an implicit community of interest in research and innovation with various backgrounds, disciplines, and areas of life at a specific locality. To make this community explicit, we address the untapped potential for professional social matching by introducing new connections with a diversification strategy that subscribes to the principle of balancing between similarity and diversity [2, 18]. Specifically, we vary degrees of diversity in the social network structures while at the same time seeking shared interests and topics by measuring the similarity of the produced content.

To apply and evaluate the recommendation strategy in practice, we collected tweets and followership data on more than 12,000 actors who follow the Twitter account of at least one of the three universities. To remedy isolated social groups on Twitter, we suggest reshaping the social network structures rather than exposing the users to more diverse content. In contrast to prior research, which typically analyzes only followership ties [19], we also use mention-based social networks since mentions are stronger interaction indicators between actors. Such an approach allows for identifying three topology-based structural positions in the active user's egocentric network [20]—*Dormant ties*, *Mention-of-Mention*, and *Community membership*. While previous research touched on three structural network positions [21], in this paper, we provide their extended definition and description of the analysis procedure and present empirical findings of an online user experiment on subjective perceptions of the produced recommendations.

To empirically study the proposed diversification strategy, we set the following research question: *How do recommendations based on the proposed structural positions associate with the subjective users' perceptions of the relevance and willingness to follow-up?* Unaware of the different recommendation groups, 39 voluntary Twitter users in the target community evaluated a total of 288 recommendations (72 from each proposed structural position and one baseline group). The analysis shows that the proposed structural positions can help introduce diversity exposure in different ways: remind about forgotten ties, motivate to connect with new people, and help enter latent communities. Thus, the paper contributes to interdisciplinary research on social and people recommender systems by proposing a nonconventional perspective for diversifying the pool of people recommendations and, prospectively, making online social networks more heterogeneous.

2. Related Work

We first outline existing conceptualizations of diversity and similarity within the context of interpersonal relationships and social matching. Next, we outline the existing people recommendation approaches and diversity-enhancing mechanisms on Twitter. Finally, we review research on user-centered evaluation of recommender systems.

2.1. Optimizing for Diversity or Similarity. Concepts of similarity and diversity are two essential polarities in social participation. Driven by the natural tendency of humans to prefer similar others [22], homogeneity is preferable when establishing trustworthy and coherent relationships [23]. At the same time, diversity is vital for productive and innovative collaboration [24]. Prior research has studied the perceived diversity of social relationships [25] and explored how diversity dimensions (e.g., cognitive, physiological, and demographic differences) are addressed in Human-Computer Interaction research [26]. User-centric recommender systems research is interested in diversity as a design goal to overcome algorithmic biases [11, 27] and drawbacks of personalization in information filtering [28, 29]. The common conceptual aspect across prior literature is that diversity is seen as the opposite of similarity [30] and has been defined, for instance, as average dissimilarity [31], distributional inequality [32], and nonredundancy [33]. Therefore, diversity can be interpreted as a perceived difference or measurable distance between all recommendations presented to the user.

While both similarity and diversity can be substantial, optimizing for either of them has been criticized [10]. For instance, social recommendations built on the principle of similarity might strengthen existing communities but can also lead to social polarization and echo chambers [34], hampering information flow, innovation, and creativity [35]. Extreme diversity among community members can negatively affect, for example, knowledge sharing and decision-making [36], resulting in conflicts, especially in the case of surface-level social and cultural differences (e.g., demographic qualities). Thus, researchers have investigated how to overcome or decrease the impact of the abovementioned adverse effects. For instance, it has been found that actors should share common ground in terms of background qualities, values, or goals to establish fruitful relationships [37]. At the same time, professional roles, capabilities, and skills should vary [38]. Rajagopal et al. [39] suggest that matching people based on dissimilarities of attitude or opinion toward the topic of interest results in better learning experiences compared to similarity-maximizing recommendation approaches. Geared toward diversity-enhancing approaches in the social matching of scholars, researchers have proposed recommending not only very similar others but also somewhat similar and different people to extend the social circles [13].

In this study, we focus on the so-called diversity exposure [10] that refers to "the content that the audience actually selects, as opposed to all the content that is available." Diversity exposure is associated with studies on detecting expertise and opinions online to extend personal autonomy (individuals' choices) and overcome echo chambers [40] for more informed rather than polarized opinions. We contribute to the research on diversity exposure by introducing a strategy that exposes Twitter users to the diversity in their social networks. Driven by the idea that fruitful relationships benefit from both shared interest and diversity, we conduct an experiment controlling for similarity and focusing on the effects of social diversity. While the concepts of diversity and similarity are well studied regarding cognitive qualities, personality, and demographics, their manifestation on social networks remains understudied. Considering that social structure encapsulates various human biases, our approach aims to decrease their impact by enhancing diversity in the composition of individuals' social networks.

2.2. Approaches to People Recommendations. Epistemologically, Twitter-based people recommendation approaches utilize user modeling based on data retrieved from basic features of the platform: "follow," "tweet," "mention," and "retweet" [19]. Accordingly, content-based approaches focus on analyzing textual content, such as tweets and retweets, while network-based approaches examine followership and mentions relationships. Table 1 provides an overview of existing approaches for recommending people on Twitter.

The most conventional approach identifies similarities in users' topics of interest content-wise and shared audience in social networks. In addition to similarity-based approaches [41], the number of followers and followees in a user profile can be used for producing recommendations based on the "popularity" dimension [42]. Recommendations can also be based on users' activities such as tweeting, mentioning, and retweeting [43]. For example, depending on the social matching scenario, the most popular and active users might be prioritized or omitted from the list of recommendations.

Since traditional recommendation approaches have been criticized for fostering human and algorithmic biases [45], recommender systems research has recently explored different approaches for diversification. In the context of people recommendations, these approaches fall into two categories. The first relates to the *diversity of features*—the most conventional approach that focuses on deriving multiple user features, that is, explicit or implicit characteristics such as interests, social network, affiliation, and others. For example, Yuan et al. [46] proposed extracting contextual features such as mobility and activity for people recommendations on Twitter. Guimarães et al. [47] proposed an extension of users' features by simultaneously utilizing content-based, collaboration-based, and user-based information, thus increasing user modeling accuracy and the effectiveness of people recommendations.

The second category relates to the *diversity of analytical procedures*—utilizing hybrid analysis techniques for filtering the recommendation pool. A representative example of this approach complements identifying the content similarity with sentiment analysis to create emotion-based recommendations for matching Twitter users with shared topics and similar [48] or different [49] emotional attitudes toward them. Jacovi et al. [50] argued for mining person and content interest relationships to complement existing approaches based on similarity and familiarity. According to the authors, merely being similar or familiar with a person does not imply directional interest, yet it is essential for establishing new ties. We also contribute to the diversity of analytical procedures by combining different analyses for social tie identification accompanied by retrieving the topic similarity from the content of users' tweets. In addition, we consider contextuality through boundary specification for the recommendation pool: geographically bounded shared interests serve as a common ground across members with inherent internal diversity of expertise sectors.

From the perspective of social networks, people recommendation approaches are limited to the analysis of followership networks and typically utilize triadic closure principles [19]. Smith et al. [51] argued that the nature and topology of networks on Twitter are underutilized, and conventional filtering or recommendation algorithms are trapping users to homogeneous content and social connections. Following the call for diversifying social network structures, Sanz-Cruzado and Castells [52] proposed recommending weak ties derived from dynamic interactive networks (e.g., based on retweets or mentions) of Twitter users. However, their evaluation study is based purely on comparing generated recommendations for forming connections in real life. Importantly, they do not ask users about their perceptions of the recommendations. Although we firmly subscribe to the overall goal of Sanz-Cruzado and Castells' work, we approach the network-based recommendation mechanisms and the evaluation procedure differently. We propose utilizing both followership networks to identify weak ties and mention-based networks to reveal interaction-driven weak and tacit connections. Aiming at user-centered evaluations beyond the accuracy [15], we also measure subjective perceptions of recommendations from different structural network positions.

2.3. User-Centric Evaluation of Recommender Systems. Recommender systems traditionally utilize system-centric evaluation methods and rarely assess the quality of recommendations with user-centric experiments [15, 53]. System-centric methods algorithmically simulate the

Table 1: Overview of typical recommendation approaches.

Approach	Type of recommended users
Similarity [41]	Users who share interests (tweet content) or audience (network)
Triadic closure [6, 19]	If A & B both follow C, recommend A to B (or vice versa); if A follows both B & C, recommend B to C; if A follows B and B follows C, recommend C to A
Popularity [42, 43]	Users who have many followers or followees
Activity [43]	Users who frequently tweet, mention, retweet, etc.
Reciprocity [44]	One's followers
Activity & followership [44]	Users whose tweets are retweeted by followees of the target user

accuracy by comparing the estimated opinions regarding the value of recommendations with pre-built ground truth datasets [54]. User-centric evaluation collects opinions and observes behavior during the interaction with the recommender system [55]. Prior research suggests that system- and user-centric measures could lead to contradictory results [56]: recommendations that the system estimate to be relevant may not be perceived the same way by the user. Therefore, the need for operationalizing subjective measures to evaluate recommendations' quality has been raised [13].

Nevertheless, the research on defining and applying subjective evaluation measures in practice remains scarce. Existing research primarily focuses on assessing the system's objective aspects, such as interaction effort and efficacy [56]. The measures that aim to reveal users' attitudes regarding recommendations are mainly driven by the idea of evaluating trust toward the system and its functional effectiveness. For instance, measures such as perceived accuracy [55] and familiarity [57] were proposed assuming that recommendations that best match the user's interests and are perceived as familiar increase the trust toward the system and imply efficiency. The subjective measures of novelty and diversity are driven by the goal of revealing the users' satisfaction [58, 59]: the novel and diverse recommendations can increase the subjectively perceived usefulness.

In summary, existing evaluation measures have been proven suitable for item recommenders (suggestions on products and content). However, as objects of recommendations, people represent more complex quality criteria that can affect decision-making regarding evaluating their value. Considering social matching scenarios for professional social networking, the subjective perceptions on recommendation relevance can be influenced by a particular need for partnering. This calls for context-specific operationalization of evaluation metrics [2]. Besides, there are no established measures for subjective perceptions of the people recommendation to our best knowledge. This paper proposes evaluating the relevance of recommendations from two perspectives—the value of recommended people for professional activities and their topics' usefulness. In addition, we operationalize measures for evaluating low- and high-cost follow-up activities.

3. Exploring and Defining Structural Network Positions

Our overall matching strategy is to introduce people who share similar interests based on the tweets' content (e.g., shared scientific interests) but have only an indirect or inactive connection in the social network. Thus, by controlling content similarity, we can compare subjective perceptions of recommendations based on the different structural positions defined as follows:

(i) *Dormant tie* (*Dorm*)—reintroducing existing followee with whom user did not have any explicit interactions; as a recommendation mechanism, this could remind about possibly ignored or forgotten ties [60];

(ii) *Mentions-of-Mentions* (*MoM*)—a friend-of-a-friend type of a connection [61] in the mention network; in contrast to typical followership networks, the mention network is based on more explicit interactions;

(iii) *Community membership* (*Com*)—a user identified to belong to the same community cluster [62]; this could introduce new people in a computationally identified network cluster with no explicit followership and mention-based ties;

(iv) *The rest* of the population (*Rest*) as a baseline condition—all the other users who follow at least one of the institutional accounts, considered as the most random source of recommendations.

In the following, we describe the procedure for exploring and defining the three structural positions as potential recommendation strategy, including details on data collection, data processing, and analysis methods.

3.1. Data Source, Cleaning, and Preprocessing. We used the official Twitter API to collect followers of to-be-merged universities and their recent tweets (See Figure 1, step 1). The raw Twitter data were stored on a MongoDB database (https://www.mongodb.com/), a flexible data model that allows development without a predefined data schema. The system is implemented in Python, and we use the PyMongo package (https://pypi.org/project/pymongo/) to set up a communication channel with a MongoDB database. We collected tweets and followership data using separate modules for each task, respectively, "GetTweets" and "GetFollowers." The preprocessing phase takes care of the tweet text cleaning task (See Figure 1, step 2) in three stages: (1) converting letters to lowercase, (2) removing the English stop words, and (3) removing the nonletter characters and URLs. Since we aimed to generate person-to-person recommendations, preprocessing also consisted of manually

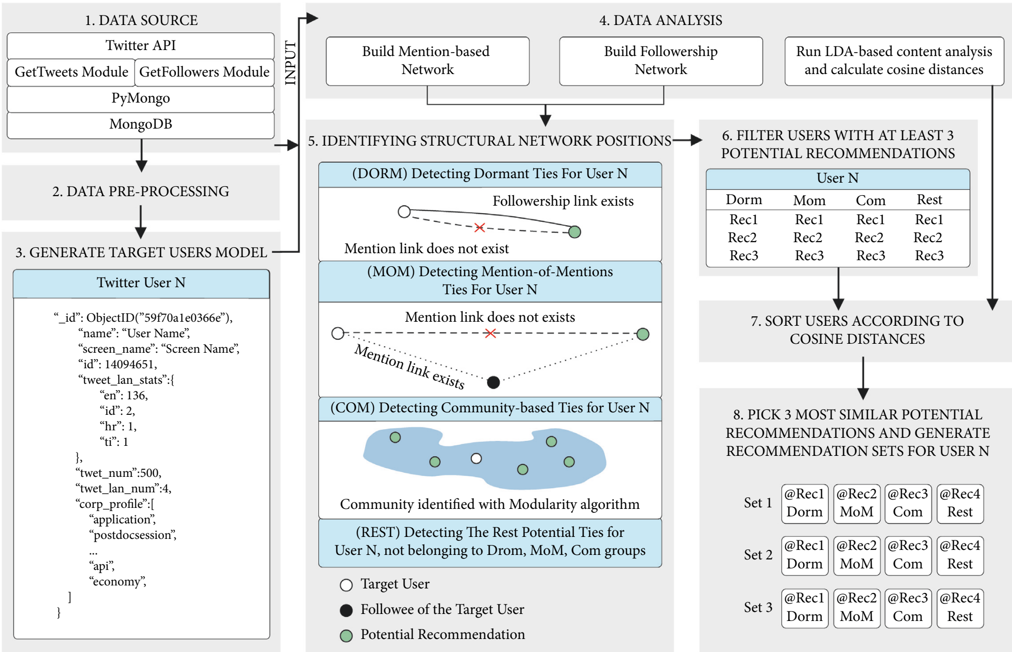


FIGURE 1: Visual representation of the data collection, analysis procedure, and recommendation approach.

filtering out the organizational accounts (e.g., Twitter profiles of local companies). Next, we generated models of target users (See Figure 1, step 3) by collecting the statistical information regarding followers of universities' accounts, including the total number of user's tweets, the number of languages in use, and the number of tweets in each language. An index is created for each user in the database. The text corpus from all the cleaned tweets of a user is collected as the corpus profile ("corp_profile").

We selected users who follow at least one of the selected four university-related Twitter accounts in Tampere, Finland. For each follower, we collected their 500 most recent tweets. The collected dataset consisted of 12,809 distinct followers and 3,523,397 tweets. The content analysis could only be done within one language corpus because of a lack of analysis procedures supporting the local language. Therefore, we excluded Twitter profiles that contain less than 50 English tweets (out of the most recent 500) from the analysis and the pool of potential recommendations. It is noteworthy that English is actively and proficiently used by most of the users in the dataset. Tweet language distribution is as follows: 58.70% are Finnish, 31.06% are English, and 10.24% are other languages. As a result, the final dataset comprises 4,474 users and 933,785 English tweets.

3.2. Social Network and Content Analyses. The data analysis (See Figure 1, step 4) comprises building mention-based and egocentric followership networks for detecting structural network positions and content analysis to obtain cosine distances for measuring the content similarity. To identify structural network positions (See Figure 1, step 5), we utilized the NetworkX Python package (https://networkx.github.io). NetworkX allows to model, manipulate, and analyze the structure of networks by specifying nodes and edges between them. In our use case, the nodes represent distinct actors in the Twitter network, and the edges illustrate the relationships between them. We created a directed mention network from the collected dataset. We used the mention network to identify the MoM and Com groups. The Mentions-of-mentions group follows the triadic closure principle—if user *A* directly mentions user *C* and *B*, then *C* would be recommended to *B* and vice versa. For the Community membership, we utilized the Louvain Modularity algorithm [63] directly in the mention-based network to detect groups of strongly related users without an explicit connection. The Dorm structural position is identified by utilizing the followership network populated directly from the Twitter API. An edge connecting two nodes in the followership network represents an existing followership link between two users on the platform. We identify the Dorm structural positions by utilizing both the mention and the followership networks.

For the experiment stage, we set the requirement for users to have at least three potential recommendations per each structural network position and apply the filter accordingly (See Figure 1, step 6).

We run the content-based similarity analysis utilizing the unsupervised topic modeling technique Latent Dirichlet

Allocation (LDA) [64]. Each topic in the LDA model is constructed with a multinomial probability distribution of words. Given *a document, in our case, a user's corpus profile, the LDA model can calculate the probabilities* of being in each topic of the document. Thus, a vector of LDA topic representation for each document can be generated, where a given number of topics $\mathbf{Z}$ within a corpus comprises documents $\mathbf{t}$. In our case, one document consists of a set of tweets per user. LDA defines each topic over a set of n-grams $\mathbf{w}$. Accordingly, the process can be formulated as follows:

$$P(w|t) = \sum_{z=1}^{Z} P(w|z)P(z|t). \quad (1)$$

Next, we use cosine distance to measure the similarity between two users as each of them has its own LDA topic vector. The lower the cosine distance value, the higher the similarity. The cosine distance is a similarity measure metric between two nonzero vectors. Given two vectors $\mathbf{u}$ and $\mathbf{v}$, the cosine distance between them is calculated as follows:

$$\mathrm{sim}(\mathbf{u}, \mathbf{v}) = \frac{\mathbf{u}^T\mathbf{v}}{\|\mathbf{u}\| \cdot \|\mathbf{v}\|}. \quad (2)$$

Therefore, content analysis (LDA + Cosine distance) allows sorting the recommendations within each group of structural positions from the most to least similar in relation to the target user (See Figure 1, step 7). Next, for each follower of the target university accounts, the top three content-wise similar users were identified within each structural position group (See Figure 1, step 8). Thus, all the given recommendations for each eligible participant have a maximum degree of similarity but belong to different structural positions. Finally, three sets of recommendations are generated for each identified eligible participant (See Figure 1, step 8). Each set consists of four recommendations: one from each of the studied structural network positions and the baseline group (Rest).

4. Experiment Design

The objective comparison of the number of possible recommendations from each group of structural positions per participant demonstrated an insignificant overlap (see Table 2). The numbers vary between users, depending on the number of followees and activity on Twitter (see examples in Figure 2). We aimed to provide a minimum of four and a maximum of twelve recommendations for each participant (1–3 from each group), which introduced the requirement for eligible respondents to have at least three other Twitter users in each structural position. 574 users out of the 4,474 met this requirement. Evaluating one set (4 recommendations) was mandatory for each participant, and the other two sets of four recommendations were voluntary. By providing up to three sets of recommendations for each participant, we wanted to achieve a higher number of evaluations per each structural network position. Some respondents evaluated all three sets of recommendations (12 in total), some only one or two sets (4 or 8 in total). This procedure resulted in subjective perceptions on 72 recommendations from each structural network position to compare them statistically.

Table 2: 26 potential recommendations from the Dorm group, 483 from MoM, and 1,246 from Com for an average targeted Twitter user. Intersections of each group are insignificant, meaning recommendation pools are mutually exclusive.

	Dorm	MoM	Com	Rest
Dorm	**26.92**			
MoM	10.66	**483.03**		
Com	0.02	0.004	**1,246.23**	
Rest	0	0	0	**8,303.3**

Bold values indicate average sizes of recommendation pool from each structural network position.

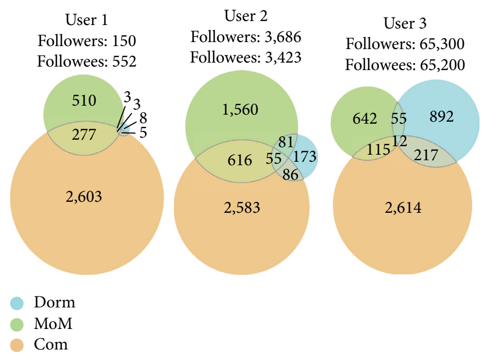


Figure 2: The sizes of recommendation pools and intersections per each structural network position for three example users with different numbers of followers and followees. The size of the Dorm group is most directly related to the number of followers/followees, while the size of MoM is affected more by activity in terms of mentions.

4.1. Procedure. The evaluation of the proposed structural network positions was carried out with two online surveys deployed on Google Forms: (1) a background questionnaire querying about demographics and the participation consent; (2) a survey with a personalized list of four other Twitter users (See Figure 3) and a set of questions to evaluate the recommendation. We chose to use Google Forms because it allows scripting-based automation to generate personalized surveys with tailored lists of recommendations.

The respondents were given no information about the types of structural network positions or why these individuals were particularly recommended to them. The order of recommendations was randomized within each set. The evaluation survey measured several subjective constructs, including perceived familiarity and perceived relevance of the recommendations and one's willingness to follow-up on them. No existing subjective measurements for perceived relevance could be found in literature, particularly for people recommendations. Therefore, the statements were operationalized based on the authors' personal experiences and insights on academic collaboration and user experience evaluation. Initially, over 20 candidate items were iteratively

Survey about potential connections on Twitter

Dear participant,

Below you can see a personalized list of potential valuable connections. First, take a look at all the profiles. When you are ready to give your opinion on each of them, please proceed to the next section. Feel free to make notes while getting familiar with their profiles. At the end of the survey, we will ask you which one of them was the most interesting to you.

Recommendation 1 First and Last Name
http://www.twitter.com/rec1_twitter_name

Recommendation 2 First and Last Name
http://www.twitter.com/rec2_twitter_name

Recommendation 3 First and Last Name
http://www.twitter.com/rec3_twitter_name

Recommendation 4 First and Last Name
http://www.twitter.com/rec4_twitter_name

Figure 3: Anonymized representation of the first set of recommendations in a survey deployed on Google forms.

assessed and refined within the project team and close colleagues, resulting in 13 items used in the survey. Perceived familiarity was measured using a 5-point Likert scale: Very unfamiliar (1)–Very familiar (5). All other items were measured using a 7-point Likert scale: Strongly Disagree (1)–Strongly Agree (7). In addition, open-ended questions inquired about the overall impressions about the recommended person and the respondent's reasoning behind the evaluations.

4.2. Recruitment and Respondents. We subscribed to the importance of research integrity and followed the policies provided by the National Ethical Committee in Finland. Accordingly, a study does not require an ethical review if it includes informed consent and does not involve any of the following: underage subjects, exposure to strong stimuli, potential long-term mental distress, or intervention with the physical integrity of participants. The study was identified as low-risk and, hence, did not require an ethical review. The participants were provided with a consent form that included a link to a detailed ethics disclaimer explaining the integrity and data management principles.

We invited eligible participants over e-mail. The targeted participants' contact information was publicly available on their Twitter profiles, and anyone could access it. The invitation consisted of a short description of the study, including links to the Background and Consent survey and a detailed ethics disclaimer. As an incentive for participation, we organized a raffle of Amazon vouchers. Out of the 574 eligible respondents, 68 signed up, and 39 participated in the experiment. They evaluated 288 recommendations in total—72 recommendations per each structural network position. The sample includes 19 male and 19 female Twitter users (one unspecified), all but one being Finnish and residents of Finland. The ages vary from Min 24 to Max 63, with an average of 43.9 and a median of 45. The majority ($N = 29$) of respondents are full-time workers, mainly university researchers. However, many other knowledge work professions are included, such as lecturers, entrepreneurs, community managers, coordinators, and project managers. The number of followers per respondent varies between 150 and 65,300, the average being 2,994 and the median 762. The number of followees varies from 311 to a maximum of 65,200, with an average of 3,289 and a median of 1254. This indicates that the respondents, on average, are active Twitter users with an extensive number of connections.

Along with the background information, we queried the respondents' typical behavior and attitudes to professional social networking with 7-point Likert statements (See Figure 4). On average, the respondents frequently network with other people, maintain their networks, and are typically careful in choosing with whom to network. From a professional perspective, their occupation typically requires intensive collaboration. Being successful in their work depends on established social ties, and they use Twitter to support their professional networks. This implies that social networking is an essential element in their professional lives. The majority also indicated that they mostly interact with like-minded people at work.

The participants were provided with a consent form that included a link to a detailed ethics disclaimer explaining the integrity and data management principles.

4.3. Survey Data Analysis. The collected responses were imported to SPSS for statistical analysis, addressing two objectives. First, we tested if the proposed structural positions are perceived to be different from each other and could thus serve as alternative analytical mechanisms. The evaluation has a thrice-repeated within-subjects design with four categorical data points per respondent. As the collected data are ordinal, we utilized a nonparametric Friedman test with Bonferroni corrected pairwise comparison to measure statistical differences. The input data for the Friedman test was in a rank format, where rank represents the frequency of each Likert scale value per evaluation statement. The second objective was to identify correlations among experimental variables. We were particularly interested in revealing whether perceived familiarity or attitude toward social and professional networking correlate with perceived relevance

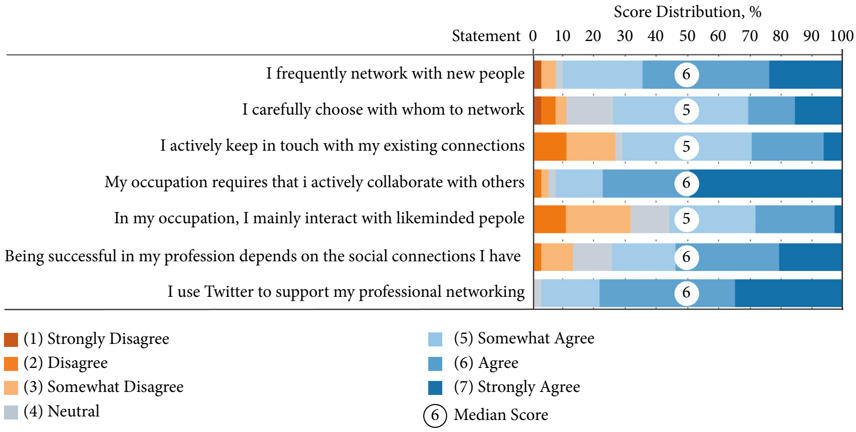


Figure 4: Overview of respondents' self-reported behavior regarding social and professional networking.

and willingness to follow-up on the recommendation. We utilized a nonparametric bi-variate Spearman correlation test as the collected data are ordinal (Likert scale).

5. Findings

We first provide results on objective measurements of similarity and respondents' perceived familiarity with recommended Twitter users, followed by subjective perceptions regarding the relevance of the recommendations. Next, we describe the respondents' readiness to engage in follow-up activities with the recommended people. Finally, we report bivariate correlation tests on associations between various variables, which provide additional insights and future research directions. In all the subsections, we first report statistical results and continue to present related qualitative findings.

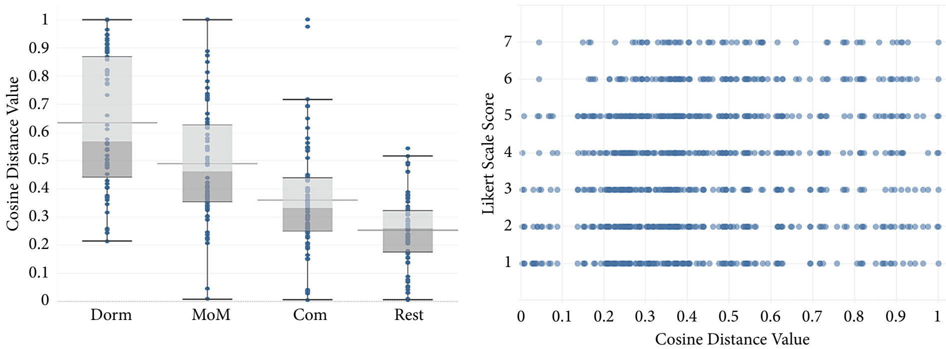
5.1. Objective Measures of Similarity. The respondents were unaware of the cosine distances (similarity measures) to avoid biased evaluations of recommendations based on the structural network positions. We aimed to pick recommendations with as equal cosine distances as possible (i.e., smallest possible variance in terms of content similarity). However, the measures are personal for each respondent and depend on the size of the recommendation pool. Since the Com and Rest recommendation pool sizes are larger, there is a higher probability of having potential recommendations with a smaller cosine distance (See Figure 5(a)). The recommendation pool of Dorm and MoM is significantly smaller, and therefore, on average, the distance is higher.

The scatterplot of respondents' scores given to recommendations over the cosine distance values demonstrates a somewhat random distribution, indicating no dependencies between them (See Figure 5(b)). A correlation test further supports this fact in Section 5.5. Therefore, we argue that the slight variance in content similarity does not prevent comparing different structural position groups.

5.2. Differences in Perceived Familiarity across the Structural Positions. The descriptive statistics results confirm that the most familiar recommendations belong to the Dorm group (See Figure 6). 44% of the recommendations from the Dorm structural network position fall into the category of either familiar or very familiar, 18% are somewhat familiar, and the remaining 38% are either unfamiliar or very unfamiliar. The other three structural groups primarily consist of unfamiliar people, with few outliers. In the MoM group, 71% of the recommended Twitter users were regarded as unfamiliar, 14% were considered familiar, and 15% somewhat familiar. The recommendations from Com and Rest groups have almost similar proportions of unfamiliar people—94% and 91%, respectively. The Friedman test indicated a statistically significant variation in respondents' ratings of perceived familiarity across different structural positions. The pairwise comparison identified substantial differences in the evaluations of the Dorm group versus other groups.

As expected, in the open-ended questions, many respondents stated that they already follow many of the recommended Twitter profiles that belong to the Dorm group. Although the respondents might be aware of the recommended person, the analysis of the social network structures revealed a lack of explicit interactions between them. The feedback in open-ended questions also supported the cases of users being unfamiliar with their followees. A relatively large number of unfamiliar followees in the Dorm group could imply that followership indeed is a weak indicator of actual social relationship and familiarity. As the act of following is typically a low-cost action, it might be even hard to keep track of and maintain their connections, especially when the number exceeds a thousand:

(Dorm) *"This person has very versatile tweets and retweets. [. . .] I already follow her, but I did not remember that."* (R25, Staff Scientist; 1,139 followers, 1,472 followees)



(a) (b)

Figure 5: (a) Distribution of cosine distance values of the evaluated recommendations from each structural network position. The lower the cosine distance value, the higher is the similarity; (b) the values of respondents' scores and cosine distances scattered relatively randomly and in similar fashion along the Likert scale (strongly agree (1)–strongly disagree (7)), which implies that there are no dependencies between the given scores and cosine distances.

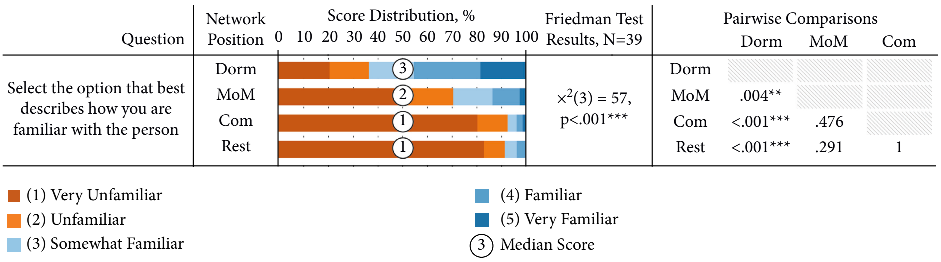


Figure 6: Proportions of perceived familiarity levels across different structural positions. 72 recommendations per position. Significance codes: 0 "***" 0.001 "**" 0.01 "*" 0.05 "" 0.1.

While the Twitter user interface allows seeing followership relationships, it is more challenging for users to reveal connections based on mentions or especially mentions-of-mentions. Only a few respondents recognized that they have a bridging tie with the recommendations from the MoM group. For instance, one noticed that they have a shared professional connection with the recommended person:

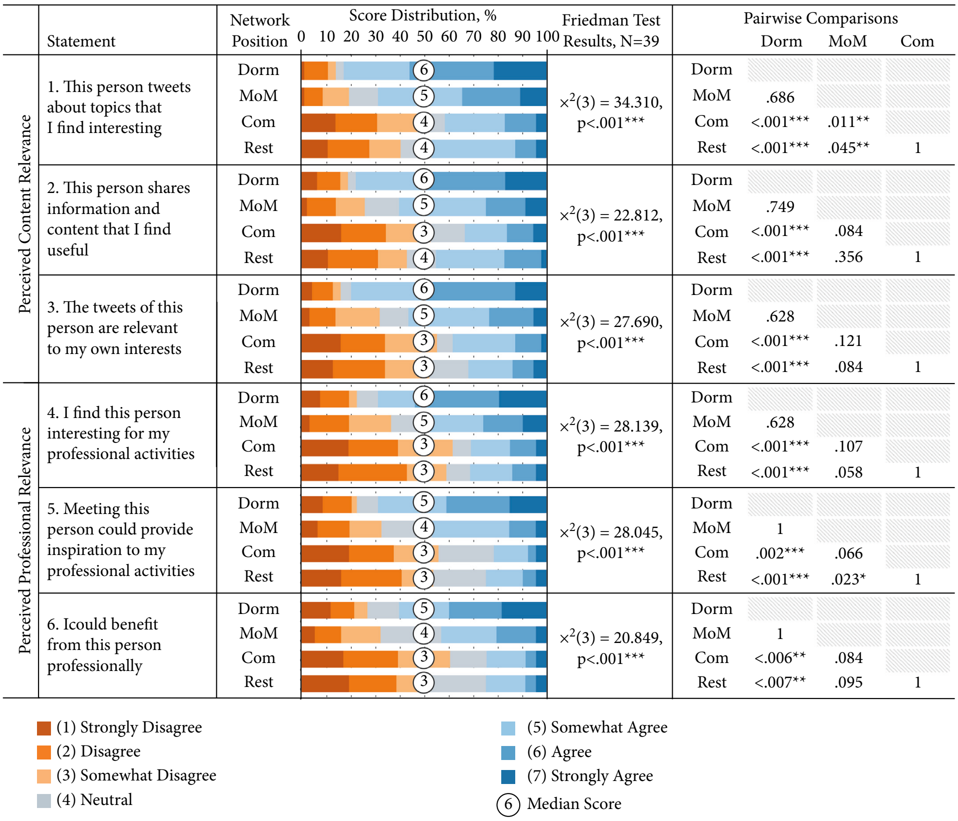
> (MoM) *"The person and her tweets are really interesting for me. She is perhaps the only one of the groups I find likely to contact and discuss future research collaboration. [...] Profile appears approachable, and she has been apparently already collaborating with some people I know."* (R1, Principal Research Scientist; 2,566 followers, 3,120 followees)

Being unaware of different structural network positions, the respondents were positively surprised by receiving many unfamiliar and diverse recommendations. In what follows, the findings demonstrate that being familiar with a person seems to increase the perception of relevance and willingness to follow-up on recommendations.

5.3. Differences in Perceived Relevance across the Structural Positions. The respondents' subjective perceptions of recommendations provided additional confirmation of the distinct nature of the three proposed structural positions. There is an apparent prevalence of positive attitude toward recommendations from the Dorm and MoM groups in the evaluations of both content relevance and professional relevance (See Figure 7). The Dorm group is perceived as the most favorable, while in the evaluations of content relevance, the opinion regarding recommendations from the Com and Rest groups split in half. Regarding the evaluation of professional relevance, the proportions of negative scores prevail.

> "*Sometimes, I am unsure whether I am answering as a professional me or a private me. For instance, when it comes to one user profile, where a private me starts to think that it is interesting due to tweets about kids, and I have kids myself.*" (R5, Project Researcher; 1,206 followers, 944 followees)

The pairwise comparison further demonstrated statistically significant differences mainly between Dorm and



	Statement	Network Position	Median Score	Friedman Test Results, N=39	Pairwise Comparisons	Dorm	MoM	Com
Perceived Content Relevance	1. This person tweets about topics that I find interesting	Dorm	6	$x^2(3) = 34.310$, p<.001***	Dorm			
		MoM	5		MoM	.686		
		Com	4		Com	<.001***	.011**	
		Rest	4		Rest	<.001***	.045**	1
	2. This person shares information and content that I find useful	Dorm	6	$x^2(3) = 22.812$, p<.001***	Dorm			
		MoM	5		MoM	.749		
		Com	3		Com	<.001***	.084	
		Rest	4		Rest	<.001***	.356	1
	3. The tweets of this person are relevant to my own interests	Dorm	6	$x^2(3) = 27.690$, p<.001***	Dorm			
		MoM	5		MoM	.628		
		Com	3		Com	<.001***	.121	
		Rest	3		Rest	<.001***	.084	1
Perceived Professional Relevance	4. I find this person interesting for my professional activities	Dorm	6	$x^2(3) = 28.139$, p<.001***	Dorm			
		MoM	5		MoM	.628		
		Com	3		Com	<.001***	.107	
		Rest	3		Rest	<.001***	.058	1
	5. Meeting this person could provide inspiration to my professional activities	Dorm	5	$x^2(3) = 28.045$, p<.001***	Dorm			
		MoM	4		MoM	1		
		Com	3		Com	.002***	.066	
		Rest	3		Rest	<.001***	.023*	1
	6. Icould benefit from this person professionally	Dorm	5	$x^2(3) = 20.849$, p<.001***	Dorm			
		MoM	4		MoM	1		
		Com	3		Com	<.006**	.084	
		Rest	3		Rest	<.007**	.095	1

FIGURE 7: Proportions of given scores regarding perceived content and professional relevance of recommendations from each structural network position, and the results of Friedman test with details on Bonferroni corrected pairwise comparison. 72 recommendations per position. Significance codes: 0~"***"~0.001~"**"~0.01~"*~0.05~"~"0.1.

Com, and Dorm and Rest. Interestingly, the difference between MoM and Com is strongly significant only in evaluating whether topics of the recommended person are of interest to the participant (Statement 1 in Figure 7).

As for the qualitative feedback, when rationalizing the relevance of recommendations, the respondents often address the importance of having similar interest topics. There is a clear positive tone in the qualitative feedback regarding recommendations from the Dorm and MoM groups. As addressed earlier, familiarity plays a significant role in evaluating relevance, and respondents often start their rationalization by explicating an existing connection with the person, if there is any. In the following example, followership relationships between the respondent and recommended person started after the face-to-face encounter at the conference:

> (Dorm) *"Lively, energetic, knows a lot about a host of topics, loves traveling. She is somebody I met at a conference a couple of years ago, and we have been in touch on social media as well."* (R3, Senior Lecturer; 295 followers, 584 followees)

In the next example, the respondent highlights that the recommended person is unfamiliar yet addresses the relevance of topics and the benefit of making a professional connection with a person from another university:

> (MoM) *"Seems active, topics relevant to me. I did not know him probably because he is in a "distant" university; it is always good to know new people from other universities."* (R13, University Researcher; 953 followers, 1,010 followees)

When evaluating recommendations from Com and Rest groups, the respondents seem to consider a variety of dimensions. For instance, the activeness of users in publishing tweets and their self-representation also play an important role in choosing whom to follow or with whom to interact:

(Com) *"Tweets very seldom; the topic is somewhat interesting, but not enough for me to follow."* (R39, HR Director; 721 followers, 1,001 followees)

(Rest) "Publisher! Seems interesting at first, but as I scrolled down, it seemed a bit too professional for my taste. The tweets in English somehow made me lose my interest." (R12, Researcher; 448 followers, 1,260 followees)

None of the respondents identified being socially connected with recommendations of the Com group, yet one respondent noticed the size of the community they share with the recommended person:

(Com) *"An interesting personality. I was able to see that we have something in common only after a closer look at the profile. Good Tweets and Retweets. [. . .] Also, the fact that we have 88 shared followers creates trust [. . .]." (R6, Project Manager; 550 followers, 763 followees)*

To sum up, the data imply that the perceptions of relevance vary across different structural positions, thus supporting the objectively observed differences between the proposed network-based matching mechanisms. Yet, relevance seems to be a weak motivator for respondents to follow-up and interact with recommended Twitter users, as discussed next.

5.4. Willingness to Follow-Up on Recommendations across the Structural Positions. The assessments of follow-up activities (See Figure 8) illustrate the weakest difference between structural positions, meaning that respondents are less open toward social interactions despite the level of recommendation relevance. Passively exploring tweets (statement 7) of recommended people from the Dorm and MoM groups is positively perceived, while other activities brought up primarily negative attitudes. The Friedman test demonstrated statistically significant differences in the evaluation of recommendations from three structural positions within all variables. The pairwise comparisons reveal apparent differences between the Dorm and Com, Dorm and Rest, MoM and Rest structural network positions regarding the intention to continue reading recommended person tweets (statement 7). There is a statistically significant difference between Dorm and Com, as well as Dorm and Rest groups regarding the attitude toward mentioning recommended people. Other activities do not illustrate very significant differences.

The respondents also addressed challenges in estimating the relevance of received recommendations and their intention for follow-up activities in the open-ended responses. For instance, one respondent mentioned that decision-making on the interestingness of a recommendation might be affected by the overall sympathy toward a Twitter user, making it challenging to draw a line between personal and professional interests:

According to one of the respondent's reasoning, as Twitter is designed mainly for distributing knowledge, it was challenging to envision social interactions with a recommended person beyond the features that the platform offers:

"*Taking a look at a person's Twitter account does not really tell me anything about what would happen if I would meet the person in real life. That is why I gave a neutral answer about the consequences of meeting with someone face-to-face. [. . .] Real-time one-to-one conversation enables quick learning and multiple ways to dig up common interests. [. . .] Twitter or other social media platforms do not provide the same opportunities; they give a very narrow view to a person, their expertise, views and what we could learn from each other.*" (R39, HR Director; 721 followers, 1,001 followees)

Even though few respondents were somewhat positive about recommendations from the baseline group (Rest), being unfamiliar with the recommended person seemed to play a role in determining the willingness to follow-up:

"*Quite general Twitter profile. Professional but also other content as well, such as news. I liked how she tweeted about the academia/academic world, although we do not work in the same field of study. I would follow her if I knew her somehow other than through Twitter. She seems nice and relatively active.*" (R12, Researcher; 448 followers, 1,260 followees).

In summary, while respondents expressed the readiness to engage in low-cost interactions, such as exploring the recommended profiles or starting to follow them, they were hesitant to consider initiating interaction beyond the Twitter platform so soon after seeing the recommendation. This is particularly the case if the actions require face-to-face interactions or direct contact. This is understandable given the short time frame for exploring the costs and benefits of potential social interaction and the limited view of the recommended person's profile. Besides, as one of the respondents admitted, Twitter is perceived as a platform for passive social behavior to broadcast and consume content. The users are accustomed to Twitter not providing features to extend the interactions to other channels.

5.5. Correlations across the Evaluation Variables. In addition to looking at how the evaluations differ across predefined structural positions, we explored various statistical associations between the variables to identify future research questions. The Spearman test revealed several statistically significant positive correlations (see Table 3). In particular, perceived familiarity positively correlates with all the other variables on subjective evaluations, especially with the perception of professional interest (statement 4) and follow-up activities such as an intention to mention the recommended Twitter user (statement 10). The respondents' background variables and social networking attitudes, such as frequency of socialization and activeness of maintaining existing ties, demonstrate a relatively strong correlation,

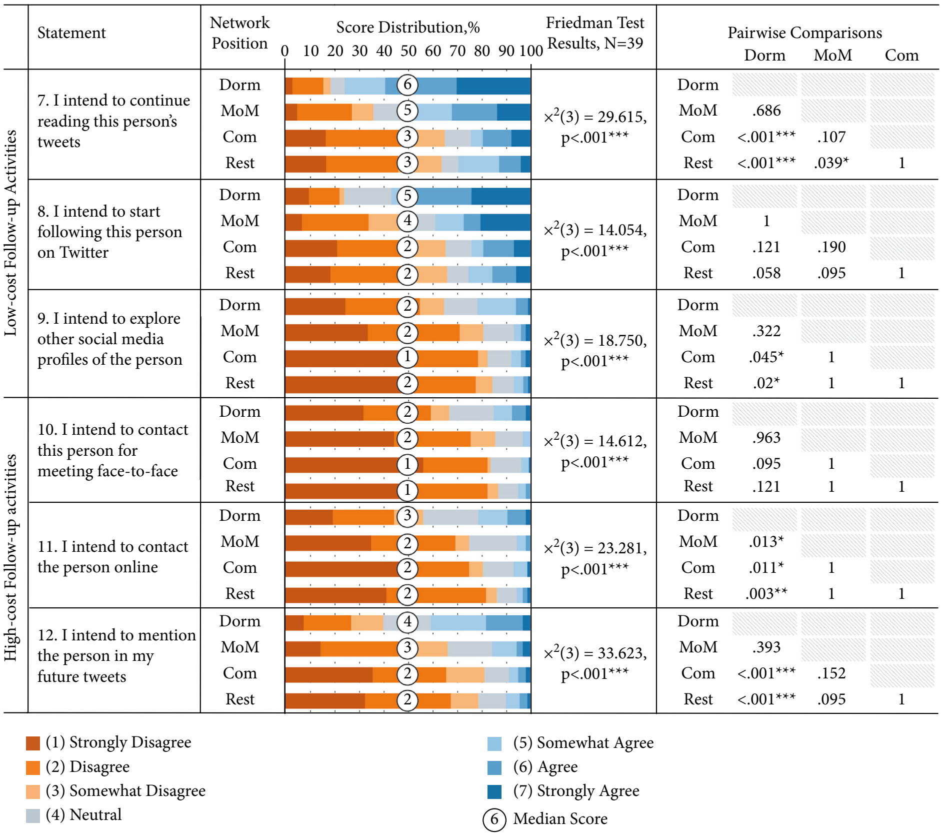


FIGURE 8: Proportions of given scores regarding willingness to follow-up on recommendations from each structural network position, and the results of Friedman test with details on Bonferroni corrected pairwise comparison. 72 recommendations per position. Significance codes: 0 "***" 0.001 "**" 0.01 "*" 0.05 "" 0.1.

particularly with the willingness to follow-up on recommendations (statements 8–12). In addition, the test indicates that the scores of high-cost follow-up activities (statements 10–12) increase with higher ratings of Twitter use for professional networking. The test results also imply no dependencies between objective measures of similarity and subjective perceptions on recommendations, which were aimed for. Overall, while correlation tests do not infer causal relations between the variables, the test hints at interesting statistical associations that should be investigated in more detail, for instance, to cover not only correlations between the recommendation evaluations but also personality and attitude-related aspects.

6. Discussion

Diversity has been a central concept in the design of information systems and social technologies, particularly CSCW and HCI research exploring its different forms (e.g., cognitive, physiological, demographic) for more inclusive and accessible technologies. This paper extended the discourse around diversity by focusing on the importance of structural diversity in social networks. We proposed identifying structural network positions in a multidimensional space of social networks, allowing the exposure of Twitter users to a variety of potential connections that they would otherwise likely miss.

Table 3: The bi-variate spearman correlation test results between variables of subjective perceptions and selected background variables, perceived familiarity, and objective measures of similarity.

	I frequently network with people		I actively keep in touch with my existing connections		I use Twitter to support my professional networking		Perceived familiarity		Objective similarity	
	Cor.	Sig.	Cor.	Sig.	Cor.	Sig.	Cor.	Sig.	Cor.	Sig.
1. This person tweets about topics that I find interesting	**0.144***	0.003	0.098	0.100	1.00	0.093	**0.479****	0.000	0.037	0.535
2. This person shares information and content that I find useful	**0.174****	0.019	**0.129***	0.030	0.104	0.080	**0.493****	0.000	0.039	0.508
3. The tweets of this person are relevant to my own interests	**0.140***	0.016	0.083	0.166	0.052	0.380	**0.437****	0.000	−0.015	0.796
4. I find this person interesting for my professional activities	**0.196****	0.001	**0.165****	0.006	**0.134***	0.024	**0.501****	0.000	0.032	0.591
5. Meeting this person could provide inspiration to my professional activities	**0.258****	0.000	**0.169****	0.005	**0.127***	0.034	**0.442****	0.000	−0.031	0.608
6. I could benefit from this person professionally	**0.199****	0.001	**0.188****	0.002	**0.156****	0.009	**0.430****	0.000	0.016	0.794
7. I intend to continue reading this person's tweets	**0.156****	0.009	**0.125***	0.035	0.111	0.062	**0.536****	0.000	0.010	0.873
8. I intend to start following this person on Twitter	**0.234****	0.000	**0.206****	0.001	**0.120***	0.047	**0.359****	0.000	0.008	0.898
9. I intend to explore other social media profiles of the person	**0.285****	0.000	**0.302****	0.000	0.107	0.073	**0.334****	0.000	0.020	0.740
10. I intend to mention the person in my future tweets	**0.302****	0.000	**0.325****	0.000	**0.265****	0.000	**0.545****	0.000	−0.024	0.686
11. I intend to contact the person online	**0.376****	0.000	**0.350****	0.000	**0.232****	0.000	**0.369****	0.000	−0.026	0.662
12. I intend to contact the person for meeting face-to-face	**0.384****	0.000	**0.346****	0.000	**0.205****	0.001	**0.270****	0.000	−0.009	0.881

Values in bold indicate correlation significance. **0.01; *0.05.

To answer our research question, the findings illustrate that recommendations are perceived differently. Thus, both the objective measurements and subjective perceptions indicate the distinct nature of the proposed three structural positions. The respondents' relatively positive evaluations of relevance suggest that the proposed recommendation strategy is a meaningful approach for diversifying people recommendations.

Furthermore, the fact that the respondents could identify relevant others from all groups implies various internal and external factors that might influence the subjective perceptions. It provides evidence that identifying recommendations within a latent community of interest (academic institutions at a specific locality in this case) is a promising approach to boundary specification.

6.1. Contributions and Reflections. Our findings contribute to the research on diversity exposure of people recommendations by defining structural positions in egocentric networks and analytical procedures for identifying related recommendations from Twitter data. The following pinpoints the possible ways of how our strategy and the different structural positions might contribute to the diversification agenda:

(i) A recommendation from the dormant ties group can remind users about their existing connections that possess expertise or perspectives that they need at the moment;

(ii) Mention-of-mention recommendations can motivate interaction with new people in a trustworthy manner as there are bridging actors in-between;

(iii) Recommendations based on community membership structural position can motivate the user to enter a latent community in a different area of the overall network, however with shared topics of interest.

Thus, the proposed structural positions can help introduce diversity exposure in different ways, prospectively suggesting connections that Twitter users would otherwise overlook.

At the same time, the findings indicate a significant role of familiarity in the subjective evaluation of recommendations. As expected, most recommended Twitter users were unfamiliar to the respondents. However, there were some familiar people, even in the "community membership" and the "rest" groups. This could be explained by the empirical context of the experiment, where boundary specification was based on the followership of the selected institutional Twitter accounts bound to a locality. We assume that this would have a strengthening effect on perceived familiarity. Research on social psychology has also revealed that homophily bias increases the perception of familiarity [65]—the stronger the perceived similarity, the more preferable and familiar the person would seem to be.

A relatively large number of unfamiliar followees in the Dorm group could imply that followership indeed is a weak indicator of actual social relationships. Thus, even though there is an established followership link, it is worthwhile to remind users about people belonging to the Dorm structural network position, which could result in more explicit social interaction. The presence of a relatively high number of outliers in the familiarity evaluation can also be explained by some respondents having a large number of followers and followees, all of whom they practically cannot remember. It has been shown that cognitive and temporal limitations prevent people from maintaining the number and quality of their relationships [66, 67]. Besides, interactions on social media might create a false sense of connection [68] and do not match with offline relationships or predict the degree of familiarity between the actors.

In addition, this study contributes to user-centric evaluation methods in the context of people recommender systems. Prior research on evaluating recommender systems is largely built on the assumption that the more accurate the algorithms, the better the user experience [15]. While this approach is useful in evaluating item recommendations (e.g., products or multimedia content), recommending people involves a different notion of recommendation quality. The operationalized measures of subjective perceptions presented in this work can be utilized in future research on evaluating the relevance and familiarity of social recommendations. However, measuring the potential follow-up activities beyond the intention to start following the recommended person worked poorly: the findings illustrate that respondents are generally not interested in speculating high-cost follow-up actions (e.g., face-to-face meetings). Thus, it is questionable to utilize such a measure as an indicator of recommendation quality, at least in the context of controlled experiments. That said, we acknowledge an inherent challenge in measuring the relevance of people recommendations. As the benefits of more heterogeneous social networks only surface over time, measuring the immediate impression about the relevance of a recommendation will likely not reflect their long-term value as a connection.

6.2. Practical Implications. Existing recommendation mechanisms shape the choices people make, influencing not only the diversity of interests and opinions but also social structures [69]. Mindful of the threat that such a high agency could strengthen structural issues like polarization and echo chambers, we believe that the recommendation strategy proposed in this article is worth pursuing in practical applications. Notably, the proposed structural network positions could contribute to systems design that can lead to more diverse exposure in individuals' social networks. The analysis procedures could be transferred to many other social media platforms; after all, the proposed content analysis is not limited to hashtags, and the notions of followership and mentions are common in other services as well.

When applying the proposed strategy and the structural network positions in a real-life people recommender system, the restrictions on eligibility criteria for the users, driven by the experiment setup, can naturally be disregarded. There might be scenarios when Twitter users do not have enough recommendations from the Dorm type of a structural position. However, our finding demonstrated that the MoM and Com groups result in numerous options even for the users with a small number of followers and followees. The effectiveness of the proposed strategy can be strengthened by increasing transparency regarding the recommendation logic in the actual system and explicating the potential value of recommendations from each group. This, in turn, might also improve the willingness to follow-up on them.

6.3. Limitations and Future Work, Experimental Setup and Generalizability. The conducted experiment naturally comes with limitations that can affect the validity and reliability of the findings. First, although the presented diversity-enhancing recommendation strategy seems promising, we could not yet compare that with other strategies in this pioneering study. Thus, the assessment of the goodness of the recommendation strategy remains quite preliminary based on this experiment. In future work, a comparison with conventional recommendation algorithms would show the goodness in relation to the currently used standards.

In addition, we only tried three calibrations within the proposed strategy and with a sample size limited by practicality and data availability. Regarding generalizability, the respondents mostly represent the same geographical area and cultural background due to the selected focus of introducing users who feel some affinity to the to-be-merged universities. The sample of participants for the experiment might also be considered biased, as the number of respondents' followees and followers is higher compared to an average Twitter user. Large-scale studies and comparisons against different baseline recommendations are required to prove the effectiveness of the overall matching strategy and the structural positions as recommendation mechanisms. Nevertheless, as the paper lays the groundwork for a new diversification strategy, it is essential to show that such an alternative strategy is sensible from the users' viewpoint and technically feasible before comparing it with others. We call for follow-up research to also compare the effectiveness of current and other alternative algorithmic approaches.

Network analysis. Our recommendation approach utilizes social network analysis, which has been critiqued due to several issues [70]. First, modeling networks is limited in terms of deriving personal roles and interpersonal experiences. It is an oxymoron to reduce multi-faceted and dynamic social relationships into network structures with simple node and edge features. Second, network-based analysis can hardly reveal a full and truthful picture of real-world relationships. Nevertheless, our study demonstrates that it is possible to analyze Twitter social networks in new ways that can advance the social matching of individuals.

Content Analysis. Due to the lack of accurate content analysis procedures for the local language, the content analysis was limited to English tweets. In addition, we did not distinguish between personally created tweets and retweeted tweets. These factors might have decreased the accuracy of representing an individual's topics of interest, which, in turn, could affect the subjective perceptions of the recommendations. This study also does not consider the segmentation of the participants according to the types of Twitter use or personality traits [71]. However, the results of correlation tests imply that such factors can be relevant. This opens avenues for investigating various personality-related and other background variables that might affect the perceived relevance of recommendations and readiness to follow-up on them.

7. Conclusion

Despite the extensive prior research on people recommendations on Twitter, the social network structure perspective has been generally underutilized. To address this gap, we proposed a new recommendation strategy for a Twitter-based implicit community of interest, prospectively producing more heterogeneous people recommendations and positively diversifying the users' social networks. The novelty of the proposed strategy lies in combining mention-based and followership-based networks to identify different types of structural network positions: dormant (followership with no explicit interactions), mention-of-mention (a friend-of-friend connection in the mention network), and community based (users belong to a shared community with no explicit followership and interactions). The findings illustrate that the proposed structural positions are indeed distinct from each other and that the respondents could find relevant users from all groups. However, the willingness to follow-up on recommendations is relatively low and primarily driven by perceived familiarity. We call for more design-oriented research to identify solutions that could increase the probability of follow-up actions. We conclude that a more comprehensive analysis of social networks and more human-centric methods to evaluate recommendations are necessary to improve the benefits and effectiveness of people recommendations on Twitter and other social networking services.

Data Availability

The data generated or analyzed during this study are included in this article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was conducted under Business Finland Project Big Match (3166/31/2017 and 3074/31/2017).

References

[1] L. Terveen and D. W. McDonald, "Social matching: a framework and research agenda," *ACM Transactions on Computer-Human Interaction*, vol. 12, no. 3, pp. 401–434, 2005.
[2] T. Olsson, J. Huhtamäki, and H. Kärkkäinen, "Directions for professional social matching systems," *Communications of the ACM*, vol. 63, no. 2, pp. 60–69, 2020.
[3] G. Kossinets and D. J. Watts, "Origins of homophily in an evolving social network," *American Journal of Sociology*, vol. 115, no. 2, pp. 405–450, 2009.
[4] S. Mizzaro, M. Pavan, and I. Scagnetto, "Content-based similarity of twitter users," in *Proceedings of the European Conference on Information Retrieval*, pp. 507–512, Vienna, Austria, March 2015.
[5] G. Carullo, A. Castiglione, A. De Santis, and F. Palmieri, "A triadic closure and homophily-based recommendation system for online social networks," *World Wide Web*, vol. 18, no. 6, pp. 1579–1601, 2015.
[6] J. Hannon, K. McCarthy, and B. Smyth, "Finding useful users on twitter: twittomender the followee recommender," in *Proceedings of the European Conference on Information Retrieval*, pp. 784–787, Dublin, Ireland, April 2011.
[7] J. Su, A. Sharma, and S. Goel, "The effect of recommendations on network structure," in *Proceedings of the 25th International Conference on World Wide Web*, pp. 1157–1167, Montreal, Canada, April 2016.
[8] C. Fuchs, *Social Media: A Critical Introduction*, Sage, London, UK, 2017.
[9] M. Kunaver and T. Požrl, "Diversity in recommender systems - a survey," *Knowledge-Based Systems*, vol. 123, pp. 154–162, 2017.
[10] N. Helberger, K. Karppinen, and L. D'Acunto, "Exposure diversity as a design principle for recommender systems," *Information, Communication & Society*, vol. 21, no. 2, pp. 191–207, 2018.
[11] B. Edizel, F. Bonchi, S. Hajian, A. Panisson, and T. Tassa, "FaiRecSys: mitigating algorithmic bias in recommender systems," *International Journal of Data Science and Analytics*, vol. 9, no. 2, pp. 197–213, 2020.
[12] E. Olshannikova, T. Olsson, J. Huhtamäki, S. Paasovaara, and H. Kärkkäinen, "From chance to serendipity: knowledge workers' experiences of serendipitous social encounters," *Advances in Human-Computer Interaction*, vol. 2020, Article ID 1827107, 18 pages, 2020.
[13] E. Olshannikova, T. Olsson, J. Huhtamäki, and P. Yao, "Scholars' perceptions of relevance in bibliography-based people recommender system," *Computer Supported Cooperative Work (CSCW)*, vol. 28, no. 3-4, pp. 357–389, 2019.
[14] A. Said and A. Bellogín, "Comparative recommender system evaluation: benchmarking recommendation frameworks," in *Proceedings of the 8th ACM Conference on Recommender Systems*, pp. 129–136, Foster City, CA, USA, October 2014.
[15] S. M. McNee, J. Riedl, and J. A. Konstan, "Being accurate is not enough: how accuracy metrics have hurt recommender systems," in *Proceedings of the CHI'06 Extended Abstracts on Human Factors in Computing Systems*, pp. 1097–1101, Montréal, Canada, April 2006.
[16] B. K. Googins and S. A. Rochlin, "Creating the partnership society: understanding the rhetoric and reality of cross-sectoral partnerships," *Business and Society Review*, vol. 105, no. 1, pp. 127–144, 2000.
[17] L. Argote and R. Ophir, "Intraorganizational learning," in *The Blackwell Companion to Organizations*, J. A. C. Baum, Ed., pp. 181–207, Blackwell Publishers Ltd., Hoboken, NJ, USA, 2002.
[18] S. Aral and M. Van Alstyne, "The diversity-bandwidth tradeoff," *American Journal of Sociology*, vol. 117, no. 1, pp. 90–171, 2011.
[19] S. M. Kywe, E.-P. Lim, and F. Zhu, "A survey of recommender systems in twitter," in *Proceedings of the International Conference on Social Informatics*, pp. 420–433, Lausanne, Switzerland, December 2012.
[20] B. L. Perry, B. A. Pescosolido, and S. P. Borgatti, *Egocentric Network Analysis: Foundations, Methods, and Models*, Vol. 44, Cambridge University Press, Cambridge, UK, 2018.
[21] E. Skenderi, E. Olshannikova, T. Olsson et al., "Investigation of egocentric social structures for diversity-enhancing followee recommendations," in *Proceedings of the ACM UMAP 2019 Adjunct: Adjunct Publication of the 27th Conference on User Modeling*, Larnaca, Cyprus, June 2019.

[22] Y. C. Yuan and G. Gay, “Homophily of network ties and bonding and bridging social capital in computer-mediated distributed teams,” *Journal of Computer-Mediated Communication*, vol. 11, no. 4, pp. 1062–1084, 2006.

[23] P. V. Marsden, “Core discussion networks of Americans,” *American Sociological Review*, vol. 52, no. 1, pp. 122–131, 1987.

[24] M. Muller, K. Ehrlich, T. Matthews, A. Perer, I. Ronen, and I. Guy, “Diversity among enterprise online communities: collaborating, teaming, and innovating through social media,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI’12*, pp. 2815–2824, ACM, Austin, TX, USA, May 2012.

[25] L. P. Robert, “Far but near or near but far?: the effects of perceived distance on the relationship between geographic dispersion and perceived diversity,” in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems CHI’16*, pp. 2461–2473, San Jose, CA, USA, May 2016.

[26] J. Himmelsbach, S. Schwarz, C. Gerdenitsch, B. Wais-Zechmann, J. Bobeth, and M. Tscheligi, “Do we care about diversity in human computer interaction: a comprehensive content analysis on diversity dimensions in research,” in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems CHI’19*, p. 490, Glasgow Scotland, UK, May 2019.

[27] C.-H. Tsai, J. Huhtamäki, T. Olsson, and P. Brusilovsky, “Diversity exposure in social recommender systems: a social capital theory perspective,” in *Proceedings of the CEUR Workshop*, vol. 2682, p. 8, Copenhagen, Denmark, March 2020.

[28] A. Koene, E. Perez, C. J. Carter et al., “Ethics of personalized information filtering,” in *Proceedings of the International Conference on Internet Science*, pp. 123–132, Brussels, Belgium, May 2015.

[29] A. Sîrbu, D. Pedreschi, F. Giannotti, and J. Kertész, “Algorithmic bias amplifies opinion fragmentation and polarization: a bounded confidence model,” *PLoS One*, vol. 14, no. 3, Article ID e0213246, 2019.

[30] B. Smyth and P. McClave, “Similarity vs. diversity,” in *Proceedings of the International Conference on Case-Based Reasoning*, pp. 347–361, Vancouver, Canada, July 2001.

[31] K. Bradley and B. Smyth, “Improving recommendation diversity,” in *Proceedings of the Twelfth Irish Conference on Artificial Intelligence and Cognitive Science*, pp. 85–94, Maynooth, Ireland, August 2001.

[32] D. M. Fleder and K. Hosanagar, “Recommender systems and their impact on sales diversity,” in *Proceedings of the 8th ACM Conference on Electronic Commerce*, pp. 192–199, San Diego, CA, USA, June 2007.

[33] S. Vargas, L. Baltrunas, A. Karatzoglou, and P. Castells, “Coverage, redundancy and size-awareness in genre diversity for recommender systems,” in *Proceedings of the 8th ACM Conference on Recommender Systems*, pp. 209–216, Foster City, CA, USA, October 2014.

[34] J. K. Lee, J. Choi, C. Kim, and Y. Kim, “Social media, network heterogeneity, and opinion polarization,” *Journal of Communication*, vol. 64, no. 4, pp. 702–722, 2014.

[35] A. Bessi, “Personality traits and echo chambers on facebook,” *Computers in Human Behavior*, vol. 65, pp. 319–324, 2016.

[36] E. V. Hobman, P. Bordia, and C. Gallois, “Perceived dissimilarity and work group involvement,” *Group & Organization Management*, vol. 29, no. 5, pp. 560–587, 2004.

[37] W. Dong, K. Ehrlich, M. M. Macy, and M. Muller, “Embracing cultural diversity: online social ties in distributed workgroups,” in *Proceedings of the ACM Conference on Computer Supported Cooperative Work and Social Computing CSCW’16*, pp. 274–287, ACM, San Francisco, CA, USA, February 2016.

[38] R. Mitchell and S. Nicholas, “Knowledge creation in groups: the value of cognitive diversity, transactive memory and open-mindedness norms,” *Electronic Journal of Knowledge Management*, vol. 4, no. 1, pp. 67–74, 2006.

[39] K. Rajagopal, J. M. Bruggen, and P. B. Sloep, “Recommending peers for learning: matching on dissimilarity in interpretations to provoke breakdown,” *British Journal of Educational Technology*, vol. 48, no. 2, pp. 385–406, 2017.

[40] Q. V. Liao and W.-T. Fu, “Expert voices in echo chambers: effects of source expertise indicators on exposure to diverse opinions,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI’14*, pp. 2745–2754, Toronto, Canada, April 2014.

[41] F. M. Rodríguez, L. M. Torres, and S. E. Garza, “Followee recommendation in Twitter using fuzzy link prediction,” *Expert Systems*, vol. 33, no. 4, pp. 349–361, 2016.

[42] R. Garcia-Gavilanes and X. Amatriain, “Weighted content based methods for recommending connections in online social networks,” in *Proceedings of the ACM Conference on Recommender Systems*, pp. 68–71, Barcelona, Spain, September 2010.

[43] J. Weng, E.-P. Lim, J. Jiang, and Q. He, “Twitterrank: finding topic-sensitive influential twitterers,” in *Proceedings of the Third ACM International Conference on Web Search and Data Mining*, pp. 261–270, New York City, NY, USA, February 2010.

[44] S. A. Golder, S. Yardi, A. Marwick, and D. Boyd, “A structural approach to contact recommendations in online social networks,” in *Proceedings of the Workshop on Search in Social Media*, pp. 1–4, Barcelona, Spain, August 2009.

[45] M. Kaminskas and D. Bridge, “Diversity, serendipity, novelty, and coverage: a survey and empirical analysis of beyond-accuracy objectives in recommender systems,” *ACM Transactions on Interactive Intelligent Systems (TiiS)*, vol. 7, no. 1, pp. 1–42, 2016.

[46] Q. Yuan, G. Cong, K. Zhao, Z. Ma, and A. Sun, “Who, where, when, and what: a nonparametric Bayesian approach to context-aware recommendation and search for twitter users,” *ACM Transactions on Information Systems (TOIS)*, vol. 33, no. 1, p. 2, 2015.

[47] S. Guimarães, M. T. Ribeiro, R. Assunção, and W. Meira Jr., “A holistic hybrid algorithm for user recommendation on twitter,” *Journal of Information and Data Management*, vol. 4, no. 3, p. 341, 2013.

[48] K. Akiyama, T. Kumamoto, and A. Nadamoto, “Emotion-based method for latent followee recommendation in Twitter,” in *Proceedings of the 19th International Conference on Information Integration and Web-Based Applications & Services*, pp. 121–125, Salzburg, Austria, December 2017.

[49] D. F. Gurini, F. Gasparetti, A. Micarelli, and G. Sansonetti, “A sentiment-based approach to twitter user recommendation,” in *Proceedings of the 5th ACM RecSys Workshop on Recommender Systems and the Social Web*, pp. 1–4, Hong Kong, China, October 2013.

[50] M. Jacovi, I. Guy, I. Ronen, A. Perer, E. Uziel, and M. Maslenko, “Digital traces of interest: deriving interest relationships from social media interactions,” in *Proceedings of the 12th European Conference on Computer Supported Cooperative Work ECSCW 2011*, pp. 21–40, Springer London, Aarhus, UK, September 2011.

[51] M. A. Smith, I. Himelboim, L. Rainie, and B. Shneiderman, “The structures of Twitter crowds and conversations,” in

Transparency in Social Media, pp. 67–108, Springer, Cham, Switzerland, 2015.

[52] J. Sanz-Cruzado and P. Castells, "Enhancing structural diversity in social networks by recommending weak ties," in *Proceedings of the 12th ACM Conference on Recommender Systems*, pp. 233–241, Vancouver, Canada, October 2018.

[53] B. P. Knijnenburg, M. C. Willemsen, and A. Kobsa, "A pragmatic procedure to support the user-centric evaluation of recommender systems," in *Proceedings of the Fifth ACM Conference on Recommender Systems*, pp. 321–324, Chicago, IL, USA, October 2011.

[54] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl, "Evaluating collaborative filtering recommender systems," *ACM Transactions on Information Systems*, vol. 22, no. 1, pp. 5–53, 2004.

[55] P. Pu, L. Chen, and R. Hu, "A user-centric evaluation framework for recommender systems," in *Proceedings of the Fifth ACM Conference on Recommender Systems*, pp. 157–164, Chicago, IL, USA, October 2011.

[56] P. Cremonesi, F. Garzotto, and R. Turrin, "User-centric vs. system-centric evaluation of recommender systems," in *Proceedings of the IFIP Conference on Human-Computer Interaction*, pp. 334–351, Cape Town, South Africa, September 2013.

[57] R. Sinha and K. Swearingen, "The role of transparency in recommender systems," in *Proceedings of the CHI'02 Extended Abstracts on Human Factors in Computing Systems*, pp. 830-831, Minneapolis, MN, USA, April 2002.

[58] P. Castells, N. J. Hurley, and S. Vargas, "Novelty and diversity in recommender systems," in *Recommender Systems Handbook*, pp. 881–918, Springer, Cham, Switzerland, 2015.

[59] T. T. Nguyen, P.-M. Hui, F. M. Harper, L. Terveen, and J. A. Konstan, "Exploring the filter bubble: the effect of using recommender systems on content diversity," in *Proceedings of the 23rd International Conference on World Wide Web*, pp. 677–686, Seoul, South Korea, April 2014.

[60] D. Z. Levin, J. Walter, and J. K. Murnighan, "Dormant ties: the value of reconnecting," *Organization Science*, vol. 22, no. 4, pp. 923–939, 2011.

[61] G. Mollenhorst, B. Völker, and H. Flap, "Shared contexts and triadic closure in core discussion networks," *Social Networks*, vol. 33, no. 4, pp. 292–302, 2011.

[62] S. Fortunato and C. Castellano, "Community structure in graphs," in *Computational Complexity: Theory, Techniques, and Applications*, R. A. Meyers, Ed., Springer, New York, NY, USA, pp. 490–512, 2012.

[63] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 10, pp. 1742–5468, 2008.

[64] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003.

[65] R. L. Moreland and R. B. Zajonc, "Exposure effects in person perception: familiarity, similarity, and attraction," *Journal of Experimental Social Psychology*, vol. 18, no. 5, pp. 395–415, 1982.

[66] R. A. Hill and R. I. M. Dunbar, "Social network size in humans," *Human Nature*, vol. 14, no. 1, pp. 53–72, 2003.

[67] T. V. Pollet, S. G. B. Roberts, and R. I. M. Dunbar, "Use of social network sites and instant messaging does not lead to increased offline social network size, or to emotionally closer relationships with offline network members," *Cyberpsychology, Behavior, and Social Networking*, vol. 14, no. 4, pp. 253–258, 2011.

[68] S. Turkle, *Alone Together: Why We Expect More from Technology and Less from Each other*, Basic Books, Hachette, UK, 2017.

[69] P. Resnick, R. K. Garrett, T. Kriplean, S. A. Munson, and N. J. Stroud, "Bursting your (filter) bubble: strategies for promoting diverse exposure," in *Proceedings of the 2013 Conference on Computer Supported Cooperative Work Companion*, pp. 95–100, San Antonio, TX, USA, February 2013.

[70] V. Khovanskaya, M. Bezaitis, and P. Sengers, "The case of the strangerationist: re-interpreting critical technical practice," in *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*, pp. 134–145, Brisbane, Australia, June 2016.

[71] D. Quercia, J. Ellis, L. Capra, and J. Crowcroft, "In the mood for being influential on twitter," in *Proceedings of the 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*, pp. 307–314, Boston, MA, USA, October 2011.