

4-24-2023

A SYSTEMATIC REVIEW OF COMPUTATIONAL METHODS IN AND RESEARCH TAXONOMY OF HOMOPHILY IN INFORMATION SYSTEMS

Yisheng LI

Toronto Metropolitan University, yisheng.li@torontomu.ca

Ravi Vatrapu

Toronto Metropolitan University, vatrapu@ryerson.ca

Morteza Zihayat

Ted Rogers School of Management, Toronto Metropolitan University, mzihayat@torontomu.ca

Follow this and additional works at: https://aisel.aisnet.org/ecis2023_rip

Recommended Citation

LI, Yisheng; Vatrapu, Ravi; and Zihayat, Morteza, "A SYSTEMATIC REVIEW OF COMPUTATIONAL METHODS IN AND RESEARCH TAXONOMY OF HOMOPHILY IN INFORMATION SYSTEMS" (2023). *ECIS 2023 Research-in-Progress Papers*. 23.

https://aisel.aisnet.org/ecis2023_rip/23

This material is brought to you by the ECIS 2023 Proceedings at AIS Electronic Library (AISeL). It has been accepted for inclusion in ECIS 2023 Research-in-Progress Papers by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

A SYSTEMATIC REVIEW OF COMPUTATIONAL METHODS IN AND RESEARCH TAXONOMY OF HOMOPHILY IN INFORMATION SYSTEMS

Research in Progress

Yisheng, Li, Centre for Digital Enterprise Analytics and Leadership, Toronto Metropolitan University, Canada, yisheng.li@torontomu.ca

Vatrapu, Ravi, Centre for Digital Enterprise Analytics and Leadership, Toronto Metropolitan University, Canada, vatrapu@torontomu.ca

Zihayat, Morteza, Centre for Digital Enterprise Analytics and Leadership, Toronto Metropolitan University, Canada, mzihayat@torontomu.ca

Abstract

Homophily is both a principle for social group formation with like-minded people as well as a mechanism for social interactions. Recent years have seen a growing body of management research on homophily particularly on large-scale social media and digital platforms. However, the predominant traditional qualitative and quantitative methods employed face validity issues and/or are not well-suited for big social data. There are scant guidelines for applying computational methods to specific research domains concerning descriptive patterns, explanatory mechanisms, or predictive indicators of homophily. To fill this research gap, this paper offers a structured review of the emerging literature on computational social science approaches to homophily with a particular emphasis on their relevance, appropriateness, and importance to information systems research. We derive a research taxonomy for homophily and offer methodological reflections and recommendations to help inform future research.

Keywords: Homophily, Computational Social Science, Social Network Analysis, Text Analytics.

1 Introduction

Homophily is a mechanism for establishing social ties with similar others (McPherson et al., 2001). The manifestations of homophily have been witnessed in not only narrow social formations such as groups, clubs, and communities (McPherson and Smith-Lovin, 1987; McPherson et al., 2001; Schaefer et al., 2011) but also broad social phenomena such as echo chambers, filter bubbles, and polarizations (Bessi et al., 2016; Cinelli et al., 2021; Kaiser and Rauchfleisch, 2020). Homophily has important managerial implications for social business in which the digitization of the business environment generates a vast amount of data tracking human sociality, values, interests, beliefs, and attitudes (Galesic et al., 2021). On the demand side, social media's participatory affordances facilitate both a bottom-up and a top-down flow of information—for instance, lay users build homophily-based trust with other similar users or expertise-based trust with experts (Jang et al., 2019; Kim, 2015). Users also garner new connections based on common interests, experiences, and beliefs, sometimes resulting in new consumer choices (Ma et al., 2015) or customer churn (Zhang et al., 2012). On the supply side, decision support systems for social media and digital platforms can leverage massive user behavior data (Yang et al., 2022) and glean actionable insights by invoking homophily theory-driven interpretations of computational results. Homophily plays a key role in electronic word-of-mouth (eWOM), viral social media marketing campaigns such as inferring which underlying topology of a social network is better for diffusion/adoption (Chang et al., 2014) and whether a brand should directly target friends of existing consumers (Song et al., 2019).

Although computational methods are increasingly being adopted for and adapted to information systems (IS) research in general and homophily research in particular, there is a dearth of comprehensive reviews and guidelines of research methodologies appropriate to homophily in digital systems. Most extant literature either qualitatively measures perceived homophily with pre-validated surveys (e.g., “My new Facebook friend: “is similar to/like me”) or quantitatively tests hypotheses about homophily with regression models (Dvir-Gvirsman, 2017; Liu and Brown, 2014; Shalizi and Thomas, 2011; Song et al., 2019). Nevertheless, traditional survey-based methods may face threats to construct validity while many regression models like logistic regression may not be a fit for some social network relationships in which observations are not independent (Chipidza and Tripp, 2018; Lazer et al., 2020). To fill the research lacuna, our paper is motivated to offer a structured review of the methodologies being leveraged to study homophily with special emphasis on state-of-the-art computational methods in IS application domains. Our review outlines a roadmap to big data-driven studies in IS by focusing on patterns, explanations, and predictions of homophily in business and social commerce settings and highlighting methodological novelty such as blending traditional research methods with machine learning (ML) that informs future research design. We propose a taxonomy that classifies research themes, data collection techniques, and computational methods in data-driven management research on the mass digitization of homophily.

The rest of this paper is organized as follows. Section 2 describes the methodological foundations drawn from the reference discipline of computational social science (CSS). Section 3 presents our literature review procedure. Section 4 reviews two CSS approaches particularly suitable for homophily—social network analysis (SNA) and text analytics. Section 5 illustrates their utility in understanding different dimensions of homophily, especially in the application domains of IS. Section 6 outlines a taxonomy and offers several methodological guidelines for future research.

2 Methodological Foundations: Computational Social Science

CSS marks a Kuhnian research paradigm shift (Chang et al., 2014) in which social scientists, behavioral scientists, cognitive scientists, computer scientists, mathematicians, physicists, and alike, have coalesced to develop and apply computational methods in conjunction with social theory and digitized information to complex, typically large-scale social and organizational phenomena (Conte et al., 2012; Lazer et al., 2020; Shah et al., 2015). CSS transforms the scope of IS research by extending the set of methodologies that enable researchers to theorize or test homophily as an explanatory mechanism in big data settings (Edelmann et al., 2020; Parameswaran and Whinston, 2007). Analytically, there are four dominant approaches in CSS (Cioffi-Revilla, 2014; Vatrappu et al., 2016)—social complexity analysis (e.g., system dynamics), social simulations (e.g., agent-based modeling), SNA (e.g., community detection), and text analytics (e.g., topic modeling). SNA and text analytics gain more predictive power by incorporating new architectures of ML such as deep learning (DL) with graphs and large language models (LLMs) that build on state-of-the-art DL architecture called transformers. Interpretable ML methods like Shapley Additive Explanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME) provide post hoc explanations of predictions made by black-box models (Behl et al., 2021; Massachs et al., 2020). CSS encompasses a spectrum of data collection techniques differing in the level of research control and offers potential triangulation (Salganik, 2019; Shah et al., 2015). At one end, observational studies repurpose two basic types of ready-made data generated by human behaviors in social networks and collected by public institutions and/or private enterprises (Galesic et al., 2021)—a) social interactions such as following, sharing, commenting, and liking, and (b) social conversations characterized by user-generated content (UGC) such as topics and sentiments (Vatrappu et al., 2016). Online users often do not tend to hide their preferences due to the affordance of pseudonymity in some platforms or the unawareness of their behaviors being observed (Salganik, 2019), which provides a window to peer into homophily in digital space. At the other end, researchers design field/lab/digital experiments to create tailor-made data for a specific research question and isolate causal effects to improve explanatory power (Hofman et al., 2021). Instead of operationalizing homophily predominantly for predictive purposes (Khanam et al., 2023), CSS can integrate descriptive, explanatory, and predictive analysis to achieve generality of findings, control of measurement, and the realism of the context simultaneously (Chang et al., 2014; Hofman et al., 2021).

3 Literature Review: Methodology

The social interaction and conversational aspect of homophily are amenable to SNA and text analytics, respectively. Therefore, the first step in our literature review procedure was formulating the search query (“homophily” AND “computational social science” OR “social network analysis” OR “text analytics” OR “natural language processing” OR “text mining”). The last two search keywords are often synonymous with text analytics. Second, we entered this query in Google Scholar, Scopus and databases like ABI/Inform, Business Source Premier, and Science Direct (within article title, abstract, and keywords). We only searched peer-reviewed articles written in English. The combination “homophily” AND “computational social science” returned rather limited results (e.g., 19 papers in Scopus). Third, we applied two key inclusion criteria—Scopus citation count (cut-off number: 10) or the Australian Business Deans’ Council (ABDC) journal quality list (ranking B and above) to ensure search quality. We mainly selected papers from IS journals, management/business journals, and high impact factor scientific publication outlets like Nature/Science/Proceedings of National Academy of Sciences, with a few exceptions of conference/journal papers that do not meet the key criteria but may demonstrate the state of the art of computational methods. We then conducted a relevance check by excluding papers that do not mention 1) the operationalization of homophily in empirical data analysis, and 2) the application of the aforementioned CSS approaches. Table 1 summarizes the key inclusion/exclusion criteria. Lastly, we backtracked the references of recent related reviews (Ertug et al., 2022; Khanam et al., 2023) to supplement our search results. Our literature review sample has $N=67$ papers. As this is a research-in-progress paper, our review is by no means exhaustive but representative and the sample will continue to expand over the course of the project.

Inclusion/Exclusion	Criteria
Inclusion: quality	Scopus citation count ≥ 10 OR ABDC Ranking $\geq B$
Inclusion: subject area	IS journals, management/business journals, high impact scientific publications
Exclusion: relevance	1) No operationalization of homophily; and 2) no application of the abovementioned computational methods

Table 1. Key Inclusion/Exclusion Criteria Summary.

4 Findings: Computational Methods for Homophily

As this paper is a methodology review, we took a concept-centric approach (Webster and Watson, 2002) to analyze the literature review sample based on computational methods, homophily-related measurements, and analytical purposes (descriptive or predictive). The majority of the papers in the sample use SNA while the rest use text analytics. Therefore, we grouped our methodological findings into two subsections—Social Network Analysis and Text Analytics. The emphasis of our analysis was synthesizing the literature to demonstrate the appropriateness and relevance of specific computational methods to homophily. We also showed the evolution of computational methods by covering several most highlighted methods in the literature (e.g., community detection, topic modeling) and some promising developments of ML like DL embeddings that are least highlighted but have the potential for making methodological contributions.

4.1 Social Network Analysis

Grounded in network theory, SNA quantifies social ties, groups, and information flows for relationship identification, prediction, or evolution (Hu et al., 2012; Kossinets and Watts, 2009; Shah et al., 2015; Susarla et al., 2012; Zhang et al., 2020). Homophily is amenable to network theory (Borgatti and Halgin, 2011) because social interactions under homophily may lead to affiliation networks represented by bipartite graphs wherein “every edge connects a node in the set of people to a node in the set of foci” (Easley and Kleinberg, 2010, p. 95). The defining structural characteristic of homophily is that the rate of contact between similar individuals is higher than the expected rate of random assortment (Şimşek and Jensen, 2008). Random graph models (e.g., Erdős–Rényi), a descriptive SNA method, can compute the expected rate as a baseline for understanding thresholds of the emergence of homophily. The small world (Watts and Strogatz, 1998) and preferential attachment (also known as scale-free or power law) (Barabási and Albert, 1999) are two canonical random graph models that can capture real-world

network features like six degrees of separation (Milgram, 1967), weak ties, clustering, and correlation. Exponential random graph model (ERGM) is one of the most highlighted and flexible models to enable statistical inferences about the influence of node attributes and structural features of homophily on network formation (Chipidza and Tripp, 2018). Node-level centrality measures are also highlighted given their relevance to homophily in diffusion (Ertug et al., 2022): 1) Closeness, the average shortest path length from a node to other nodes, suggests the ease/speed to spread something like influence sequentially; 2) Betweenness, the number of shortest paths that pass through a given node, indicates the probability of a node bridging disparate groups of nodes; 3) Eigenvector is a more sophisticated measure of influential nodes (i.e., authorities) and their directly connected nodes (i.e., hubs) (Lynn et al., 2020). Several common network-level measures capture homophily: 1) Modularity is the difference between the fraction of edges within groups and the expected fraction in a randomly wired network (Vaanunu and Avin, 2018); 2) Assortativity coefficient is the Pearson correlation coefficient of pairs of linked nodes with respect to an attribute like gender (Bucur, 2019); 3) The Silhouette value measures how similar a user is to the rest of the users in the same cluster (Kafeza et al., 2021); 4) Network density is a result of the connectedness within the clusters (Himmelboim et al., 2017); 5) Cluster coefficient is a result of the connectedness of a node's neighbors (Laniado et al., 2016). As the most highlighted descriptive SNA method to identify homophilous communities, traditional community detection algorithms partition a social network into non-overlapping groups of nodes such that there are more edges within groups than between them and groups are often sparsely connected (Girvan and Newman, 2002). For instance, optimization-based algorithms like Louvain and Markov Chain Monte Carlo (MCMC) (Handcock et al., 2007) maximize modularity (Fortunato and Newman, 2022), whereas random walk-based algorithms like InfoMap are inspired by probability and information theory. Louvain is highly popular and computationally efficient for mining large-scale social networks (Bucur, 2019). A new trend that is less highlighted in the management/business literature is moving beyond the prevalent descriptive SNA of homophily to the predictive/ML approach like graph-based DL embeddings that can calculate attribute similarity and detect overlapping communities (Fortunato and Newman, 2022). For example, Node2Vec can better preserve homophily in a network and convert multiple attributes of a node to a vector of numerical values equivalent to coordinates that pinpoint the node in the network where edges tend to appear between nearby nodes (Goyal and Ferrara, 2018).

4.2 Text Analytics

The majority of extant computational studies of homophily have been focused on SNA of structural properties (e.g., centrality, modularity) that concern how actors structurally relate to each other and organize, with limited attention dedicated to the proliferation of UGC in terms of ideas, aspirations, values, and identities that socially motivate and influence online communities (Vatrapu et al., 2016). Text analytics can discover semantic homophily (Šćepanović et al., 2017) in terms of similar linguistic features, shared purposes/experience (Wang et al., 2019b), and cultural orientations in rating and reviews (Shah et al., 2015). As one of the most highlighted rule-based, knowledge-heavy, descriptive text analytics methods, Linguistic Inquiry and Word Count (LIWC) developed by Tausczik and Pennebaker (2010) has pre-validated dictionaries to count words in psychologically meaningful categories such as emotionality, social relationships, and thinking styles, which can be used to measure linguistic similarities (i.e., linguistic homophily) in online user reviews (Kovacs and Kleinbaum, 2020). Aspect-based sentiment analysis recognizes all sentiments within a customer review, maps each sentiment to the corresponding aspect of the review, and helps identify homophily in cross-cultural settings when Asian and non-Asian guests complain about hotel service failures (Sann and Lai, 2020). Sentiment polarity analysis can detect gender homophily in high-stakes entrepreneurial pitching (Khurana and Lee, 2023).

A stochastic, knowledge-light ML approach becomes dominant in text analytics (Hannigan et al., 2019). Supervised learning (e.g., Support Vector Machine) coupled with engineered n-grams and sentiment-based features can predict stance homophily in online debates (Lai et al., 2019) and exploit homophily for link prediction in social networks (Yuan et al., 2014), while unsupervised learning (e.g., topic modeling) helps predict whether social media users posting similar topics are connected (Khanam et al., 2023). As another highlighted descriptive text analytics method, topic modeling excels in juxtaposing data and theory, understanding opinion dynamics, developing inductive text classification

systems, and creating new theoretical artifacts such as emerging themes and the links between them (Hannigan et al., 2019; Kapoor et al., 2018). Latent Dirichlet Allocation (LDA) remains an effective topic modeling algorithm in the era of DL. Topic modeling obtains a two-dimensional map showing the proximity of words based on their co-occurrence patterns and frequently co-occurring words are likely to be related and induce a topic latent in the corpus (Kapoor et al., 2018). Its unsupervised nature makes topic modeling a flexible means to discover the sense of community in content since no labelled text for training or a prespecified number of topics is required. A trend underexploited by the management/business literature is moving beyond traditional syntax-free topic modeling algorithms like LDA to more context-aware word embeddings (Evans and Aceves, 2016). An early approach to word embeddings is Latent Semantic Analysis of topological positions of words that can measure moral purity homophily (i.e., moral rhetoric) in tweets (Dehghani et al., 2016). DL embeddings like Word2Vec represent each word in the document as a vector of coordinates such that words sharing many contexts are positioned nearby (Kozłowski et al., 2019) and better learn “semantic compositionality within texts that then can be used for explanatory and predictive purposes” (Hannigan et al., 2019, p. 608). The latest LLMs like Bidirectional Encoder Representations from Transformers (BERT) are pre-trained on massive open-source corpora (e.g., Wikipedia) and fine-tuned on domain-specific datasets like Reddit posts on vaccines (Li et al., 2022), require minimum data pre-processing, and outperform predecessors on a wide range of tasks such as sentiment analysis and topic modeling.

5 Findings: Homophily in Information Systems Domains

In this section, the emphasis of our analysis was situating the aforementioned SNA and text analytics methods and the operationalization of homophily in three IS-related application domains represented by three subsections that report our domain-specific findings. Recent relevant reviews (Ertug et al., 2022; Khanam et al., 2023) informed the curation of the first two domains (i.e., subtitles 5.1 and 5.2) that are highlighted in the IS literature, whereas subtitle 5.3 is an emerging domain in the CSS literature. In each subsection, we again took a concept-centric approach to briefly summarize key concepts including the existing data collection techniques, computational methods used by the extant literature, and established/emerging themes for IS research.

5.1 Link (Friendship, Trust) Prediction

One major IS-related application domain of homophily is link (e.g., friendship, trust) prediction using topological features or/and topical components of observational social media data to predict the emergence of online relationships (Aiello et al., 2012; Khanam et al., 2023; Tang et al., 2013). The most highlighted IS research theme under this domain is homophily in location-sharing applications (e.g., Foursquare) that primarily conducts SNA of geospatial data sometimes with text analytics providing specific contexts. A study constructs dyadic user similarity values and proximity measures by applying LDA to user-generated biographies and their associated tweets and mining check-in records (Lee et al., 2016). Its instrumental variable design and counterfactual analysis of the number of formed links without homophily add explanatory power to the finding on location-based homophily in network formation, which has implications for restaurants’ seeding strategies (i.e., marketing a new product by distributing free samples). Another study proposes an ML framework that harnesses stochastic gradient descent and matrix factorization to predict the number of check-ins between user-venue pairs (Doan and Lim, 2019). The results show incorporating homophily improves prediction accuracy and could help tackle the “cold start” problem in recommender systems when new users have very few check-ins. Digital MyPersonality survey, a data collection technique that has higher research control than observational data, introduces the personality dimensions (e.g., openness, agreeableness) of homophily to explain check-in patterns (Noë et al., 2016). A fine-grained location-based homophily study predicts which venue a tweet is posted from (Chong and Lim, 2018). Privacy (e.g., check-ins) may reach untrusted recipients through the resharing affordances of social media, which warrants a nascent research theme called privacy and trust prediction. Instrumental in this theme is a paper proposing a community detection method for access control prediction and classifying social connections as trusted or untrusted (Ferreira et al., 2022). The authors use simulated data to run digital experiments in which they can exert high research control and test different homophily conditions (i.e., network configurations) in the absence of real-world data. BERT-based sentiment analysis demonstrates its

utility of privacy leakage identification on Twitter (Mittal et al., 2020). A DL model of homophily integrates user reviews, ratings, and item characteristics to create latent user representations by Doc2Vec (akin to Word2Vec) and predicts trust relations by calculating user feature vector cosine similarity (Wang et al., 2019a). Another emerging theme is music homophily in online social networks. A study develops listening similarity measures and uses K-means to cluster users based on music preferences (e.g., positive, energetic) and social attributes like gender (Zhou et al., 2018). A recent paper develops information entropy-based measures of user preferences such as mainstream, novel, or diverse music and employs a computationally efficient ML algorithm (XGBoost) with network features to predict friendship, which informs socially context-aware music recommendations for new users (Duricic et al., 2021).

5.2 Diffusion, eWOM, and eWOB

Another major IS-related application domain of homophily is diffusion in which the most highlighted research theme aims to separate homophily and social contagion (i.e., social influence) in observational social network data by means of ML and research methods like randomized digital experiments without surveying individuals about homophily's impact on eWOM (Chang et al., 2014; Chen et al., 2019; Qiu et al., 2018; Zhang et al., 2018; Zhang and Pelechris, 2014). A large-scale social network simulation study finds that seeding strategies are overestimated without considering homophily but are more effective under greater social contagion (Aral et al., 2013). A kernel-based ML approach to social recommender systems integrating homophily and contagion is more accurate than collaborative filtering (Li et al., 2014), whereas novel recommendation frameworks utilize unsupervised DL with graphs to capture homophily in both the user and item space based on co-occurrence and feature similarity (Narang et al., 2021) or predict online content popularity in the early trajectory of diffusion (Shang et al., 2022). An MCMC simulation tracks the co-evolution of social networks and the production of UGC, disentangles homophily and contagion, and enables predictive models of diffusion (Bhattacharya et al., 2019). A methodologically novel study uses LIWC to measure personality similarity and LDA to compute topical similarity that reveals latent common interests, controlling homophily by a DL graph model called DeepWalk that learns latent user representations and their similarity (Adamopoulos et al., 2018). Its quasi-experiment design leverages the affordance of Twitter—the absolute position of the “@” character in a tweet—to control content visibility to other users. The authors find 1) introverted users are more responsive to WOM; 2) agreeable, conscientious, and open-mind users are better messengers of WOM. The ensuing study computes user similarity in terms of topics in 140 million eWOM messages generated by LDA and Jaccard coefficient of followers and followees on Twitter and employs DL-based sentiment analysis of the messages, controlling latent user homophily by incorporating user characteristics and embeddings based on DeepWalk and Node2Vec in an econometric model (Todri et al., 2022). Applying LDA and controlling exposure bias caused by content homophily, a study ascribes consistent topics in tweets to needs for maintaining online persona and personal brand strategies (Geva et al., 2019). A second research theme is moving beyond homophily and health behavior in offline social networks (Christakis and Fowler, 2007) to e-health settings. Utilizing ERGM and sentiment analysis, Liu et al. (2020) build a reply network of a 10-year dataset of online healthcare communities to identify social support and reciprocal information exchange patterns. They find homophily vis-à-vis expertise such that homophily manifests in general conversations but is weakened in knowledge exchange. This new line of research could benefit from recent development in electronic Word of Behavior (eWOB) using automated capture, analysis and reporting of users' implicit online behaviours instead of their explicit ratings or reviews or opinions (Kunst et al., 2022; Kunst and Vatrappu, 2019).

5.3 Echo Chambers

IS research on echo chambers and misinformation (Markgraf and Schoch, 2019; Schuetz et al., 2021) can build upon the analytics frameworks (SNA, text analytics, and ML) and operational definitions of opinion-based homophily from a growing body of CSS literature that uncovers ideological alignment and identifies the stance from observational user behavior data (ALDayel and Magdy, 2021; Chang et al., 2014; Kaiser and Rauchfleisch, 2020; Wanek and Hidalgo, 2022; Williams et al., 2015). A communication study applies a chain of ML tweet classifiers, first classifying political/non-political content and subsequently democrat/republican users, followed by an SNA of homophily with random

graphs as the baseline (Colleoni et al., 2014). A new study performs a cross-platform analysis of 100 million social media posts to operationalize echo chambers by measuring users' leaning for controversial topics like abortion and gun control and discovers robust homophily on Facebook and Twitter but weak homophily on Reddit by implementing Louvain (Cinelli et al., 2021). A study finds that Twitter reply networks in some political debates exhibit inverse homophily by stance and lower modularity values, which implies Twitter reply-to interactions are a suitable structure to investigate how users express diverging opinions and increase cross-stance interactions (Lai et al., 2019). An ERGM of a Twitter mention network discovers homophily in ideological hashtags that signal identity or stance instead of conceptual hashtags that are used as personal references (Xu and Zhou, 2020). A methodologically novel study uses text analytics like sentiment analysis and ML to predict homophily in political support on Reddit, explaining the key predictors of user participation according to SHAP (Massachs et al., 2020). A digital governance study constructs a semantic network of tweets (nodes being words and edges being relationships between words) and uses Louvain to discover users preferentially retweeting other users with similar opinions (i.e., homophily) in e-petition discussions (Asher et al., 2019). Wang and Song (2020) combine SNA and qualitative content analysis to measure homophily in misinformation on genetically modified organisms by assortativity coefficient. Li et al. (2022) train a BERT-based classifier to predict Reddit user stance (pro or anti-vaccine) and obtain each post/comment's analytical thinking score by LIWC, examining homophily in vaccine hesitancy.

6 A Research Taxonomy of Homophily for Information Systems

We synthesized our literature review findings in a morphological taxonomy (Table 2) as this is a common practice (Gelhaar et al., 2021). Taxonomies are a suitable classification scheme of the dimensions and characteristics of application-oriented objects by summarizing the commonalities found in the literature (Gregor, 2006; Nickerson et al., 2013). We mainly took the conceptual-to-empirical approach (Nickerson et al., 2013) that used the literature as a source for deductively deriving the characteristics and dimensions of the taxonomy. The theme dimension lists both established and emerging themes under IS application domains of homophily. Themes are supported by key citations. The data collection technique dimension is a summary of different data types for evaluation and validation in the literature in terms of the level of research control. The analytical purpose dimension maps a study to a descriptive/explanatory/predictive purpose (Shmueli and Koppius, 2011). We derived the two preceding dimensions and their characteristics from CSS as methodological foundations. The computational method dimension documents specific SNA, text analytics, and ML techniques in the literature. Ideally, the characteristics of a dimension should be mutually exclusive (Nickerson et al., 2013), but in practice researchers sometimes have to trade off exclusivity for the parsimoniousness of a taxonomy (Hunke et al., 2019; Möller et al., 2020). A study in our literature review may have different data collection techniques, computational methods, and analytical purposes. Our taxonomy helps IS researchers determine where a research problem fits with existing computational methods, data collection techniques, and research themes to know if it is a significant departure from or just a variation of the extant data-driven management research on homophily.

We offer several methodological reflections and recommendations for future IS research concerning homophily. *First*, we recommend conducting cross-platform analyses like Cinelli et al. (2021) in Section 5.3. Homophily does not always “naturally” take place on social media but rather is often engineered or mediated by recommender algorithms. The interplay of affordances entwining users and news feed algorithms vary from platform to platform (e.g., discussion forum-like Reddit vs. social network-like Facebook and micro-blog like Twitter) and recommender algorithms are more configurable by users in some platforms but are not in others (De Francisci Morales et al., 2021). A cross-platform analysis would better account for the difference in the strength of homophily that is partially induced by algorithms (Salganik, 2019). *Second*, from a data collection technique standpoint, we recommend matching big observational data to other data sources that may help establish ground truth (e.g., survey). Recent studies merge administrative, biographical, and publication data by deterministic and probabilistic record linkage techniques to detect gender homophily in research collaboration networks (Kwiek and Roszka, 2021) and link homophily in massive Facebook friendships with Census to quantify social capital (Chetty et al., 2022). *Third*, from a computational methods perspective, we recommend experimenting with novel techniques like DL embeddings to detect

homophilous communities that intersect with one another in addition to traditional community detection algorithms like Louvain. It would be valuable to have a comparative analysis of topic modeling techniques including LDA and more nuanced LLMs like dynamic topic modeling with BERT. *Fourth and Last*, as a few studies in our review demonstrate (Adamopoulos et al., 2018; Kafeza et al., 2021; Lee et al., 2016; Massachs et al., 2020), a path forward is combining SNA and text analytics and leveraging interpretable ML to extract key predictors of homophily. It is worth attempting causal inference in SNA of homophily by unifying ML, experiment design, and econometric models. The next step of our research is following the guidelines (Nickerson et al., 2013; Webster and Watson, 2002) to refine the proposed baseline taxonomy and expand this paper to a full systematic review. Our findings of location-based and music homophily in Section 5.1 have the potential to be applied in our new research project that aims to investigate spatial and cultural homophily in a world-renowned music festival using privacy-preserving tracking data and music metadata.

Dimension	Characteristics						
Application Domain	Link (Friendship, Trust) Prediction		Diffusion & eWOM & eWOB		Echo Chamber		
Research Theme	Location-Based Homophily (Lee et al., 2016)	Privacy & Trust Prediction (Ferreyra et al., 2022)	Music Homophily (Zhou et al., 2018)	Homophily vs. Contagion (Adamopoulos et al., 2018)	Homophily in E-Health (Liu et al., 2020)	Opinion-Based Homophily (Xu and Zhou 2020)	
CSS Approach	Social Network Analysis (SNA)			Text Analytics			
Data Collection Technique	Observational (low research control)		Survey (medium research control)		Experiment (high research control)		
Analytical Purpose	Descriptive		Explanatory		Predictive		
Computational Method	ERGM	Louvain	MCMC Simulation	K-means	DeepWalk	Node2Vec	LIWC
	LDA	Doc2Vec	Aspect-Based Sentiment analysis	BERT-based sentiment analysis	XGBoost	SHAP	

Table 2. A Taxonomy of Homophily for IS Research.

References

- Adamopoulos, P., Ghose, A., and Todri, V. (2018). "The Impact of User Personality Traits on Word of Mouth: Text-Mining Social Media Platforms," *Information Systems Research* 29 (3), 612-640.
- Aiello, L. M., Barrat, A., Schifanella, R., Cattuto, C., Markines, B., and Menczer, F. (2012). "Friendship Prediction and Homophily in Social Media," *ACM Transactions on the Web* 6 (2), 1-33.
- ALDayel, A., and Magdy, W. (2021). "Stance Detection on Social Media: State of the Art and Trends," *Information Processing & Management* 58 (4), 102597.
- Aral, S., Muchnik, L., and Sundararajan, A. (2013). "Engineering Social Contagions: Optimal Network Seeding in the Presence of Homophily," *Network Science* 1 (2), 125-153.
- Asher, M., Leston-Bandeira, C., and Spaiser, V. (2019). "Do Parliamentary Debates of E-Petitions Enhance Public Engagement with Parliament? An Analysis of Twitter Conversations," *Policy & Internet* 11 (2), 149-171.
- Barabási, A.-L., and Albert, R. (1999). "Emergence of Scaling in Random Networks," *Science* 286 (5439), 509-512.
- Behl, S., Rao, A., Aggarwal, S., Chadha, S., and Pannu, H. (2021). "Twitter for Disaster Relief through Sentiment Analysis for Covid-19 and Natural Hazard Crises," *International Journal of Disaster Risk Reduction* 55, 102101.

- Bessi, A., Petroni, F., Vicario, M. D., Zollo, F., Anagnostopoulos, A., Scala, A., Caldarelli, G., and Quattrociocchi, W. (2016). "Homophily and Polarization in the Age of Misinformation," *The European Physical Journal Special Topics* 225 (10), 2047-2059.
- Bhattacharya, P., Phan, T. Q., Bai, X., and Airolidi, E. M. (2019). "A Coevolution Model of Network Structure and User Behavior: The Case of Content Generation in Online Social Networks," *Information Systems Research* 30 (1), 117-132.
- Borgatti, S. P., and Halgin, D. S. (2011). "On Network Theory," *Organization Science* 22 (5), 1168-1181.
- Bucur, D. (2019). "Gender Homophily in Online Book Networks," *Information Sciences* 481, 229-243.
- Chang, R. M., Kauffman, R. J., and Kwon, Y. (2014). "Understanding the Paradigm Shift to Computational Social Science in the Presence of Big Data," *Decision Support Systems* 63, 67-80.
- Chen, Z.-Y., Fan, Z.-P., and Sun, M. (2019). "Individual-Level Social Influence Identification in Social Media: A Learning-Simulation Coordinated Method," *European Journal of Operational Research* 273 (3), 1005-1015.
- Chetty, R., Jackson, M. O., Kuchler, T., Stroebe, J., Hendren, N., Fluegge, R. B., Gong, S., Gonzalez, F., Grondin, A., and Jacob, M. (2022). "Social Capital I: Measurement and Associations with Economic Mobility," *Nature* 608 (7921), 108-121.
- Chipidza, W., and Tripp, J. F. (2018). "The Social Structure of the Information Systems Collaboration Network: Centers of Influence and Antecedents of Tie Formation," *Communications of the Association for Information Systems* 42 (1), 16.
- Chong, W.-H., and Lim, E.-P. (2018). "Exploiting User and Venue Characteristics for Fine-Grained Tweet Geolocation," *ACM Transactions on Information Systems* 36 (3), 1-34.
- Christakis, N. A., and Fowler, J. H. (2007). "The Spread of Obesity in a Large Social Network over 32 Years," *New England Journal of Medicine* 357 (4), 370-379.
- Cinelli, M., Morales, G. D. F., Galeazzi, A., Quattrociocchi, W., and Starnini, M. (2021). "The Echo Chamber Effect on Social Media," *Proceedings of the National Academy of Sciences* 118 (9), e2023301118.
- Cioffi-Revilla, C. (2014). "Introduction to Computational Social Science," *London and Heidelberg: Springer*, 464.
- Colleoni, E., Rozza, A., and Arvidsson, A. (2014). "Echo Chamber or Public Sphere? Predicting Political Orientation and Measuring Political Homophily in Twitter Using Big Data," *Journal of Communication* 64 (2), 317-332.
- Conte, R., Gilbert, N., Bonelli, G., Cioffi-Revilla, C., Deffuant, G., Kertesz, J., Loreto, V., Moat, S., Nadal, J.-P., and Sanchez, A. (2012). "Manifesto of Computational Social Science," *The European Physical Journal Special Topics* 214 (1), 325-346.
- De Francisci Morales, G., Monti, C., and Starnini, M. (2021). "No Echo in the Chambers of Political Interactions on Reddit," *Scientific Reports* 11 (1), 1-12.
- Dehghani, M., Johnson, K., Hoover, J., Sagi, E., Garten, J., Parmar, N. J., Vaisey, S., Iliev, R., and Graham, J. (2016). "Purity Homophily in Social Networks," *Journal of Experimental Psychology: General* 145 (3), 366.
- Doan, T.-N., and Lim, E.-P. (2019). "Modeling Location-Based Social Network Data with Area Attraction and Neighborhood Competition," *Data Mining and Knowledge Discovery* 33 (1), 58-95.
- Duricic, T., Kowald, D., Schedl, M., and Lex, E. (2021). "My Friends Also Prefer Diverse Music: Homophily and Link Prediction with User Preferences for Mainstream, Novelty, and Diversity in Music," *Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, Virtual Event Netherlands, 447-454.
- Dvir-Gvirsman, S. (2017). "Media Audience Homophily: Partisan Websites, Audience Identity and Polarization Processes," *New Media & Society* 19 (7), 1072-1091.
- Easley, D., and Kleinberg, J. (2010). *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*, Cambridge: Cambridge University Press.
- Edelmann, A., Wolff, T., Montagne, D., and Bail, C. A. (2020). "Computational Social Science and Sociology," *Annual Review of Sociology* 46 (1), 61.
- Ertug, G., Brennecke, J., Kovács, B., and Zou, T. (2022). "What Does Homophily Do? A Review of the Consequences of Homophily," *Academy of Management Annals* 16 (1), 38-69.

- Evans, J. A., and Aceves, P. (2016). "Machine Translation: Mining Text for Social Theory," *Annual Review of Sociology* 42, 21-50.
- Ferreira, N. E. D., Hecking, T., Aïmeur, E., Heisel, M., and Hoppe, H. U. (2022). "Community Detection for Access-Control Decisions: Analysing the Role of Homophily and Information Diffusion in Online Social Networks," *Online Social Networks and Media* 29, 100203.
- Fortunato, S., and Newman, M. E. (2022). "20 Years of Network Community Detection," *Nature Physics* 18 (8), 848-850.
- Galesic, M., Bruine de Bruin, W., Dalege, J., Feld, S. L., Kreuter, F., Olsson, H., Prelec, D., Stein, D. L., and van Der Does, T. (2021). "Human Social Sensing Is an Untapped Resource for Computational Social Science," *Nature* 595 (7866), 214-222.
- Gelhaar, J., Groß, T., and Otto, B. (2021). "A Taxonomy for Data Ecosystems." *Proceedings of the 54th Hawaii International Conference on System Sciences*, Waikoloa Village, Hawaii, USA.
- Geva, H., Oestreicher-Singer, G., and Saar-Tsechansky, M. (2019). "Using Retweets When Shaping Our Online Persona: Topic Modeling Approach," *MIS Quarterly* 43 (2), 501-524.
- Girvan, M., and Newman, M. E. (2002). "Community Structure in Social and Biological Networks," *Proceedings of the National Academy of Sciences* 99 (12), 7821-7826.
- Goyal, P., and Ferrara, E. (2018). "Graph Embedding Techniques, Applications, and Performance: A Survey," *Knowledge-Based Systems* 151, 78-94.
- Gregor, S. (2006). "The Nature of Theory in Information Systems," *MIS Quarterly* 30 (3), 611-642.
- Handcock, M. S., Raftery, A. E., and Tantrum, J. M. (2007). "Model-Based Clustering for Social Networks," *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 170 (2), 301-354.
- Hannigan, T. R., Haans, R. F., Vakili, K., Tchaljian, H., Glaser, V. L., Wang, M. S., Kaplan, S., and Jennings, P. D. (2019). "Topic Modeling in Management Research: Rendering New Theory from Textual Data," *Academy of Management Annals* 13 (2), 586-632.
- Himelboim, I., Smith, M. A., Rainie, L., Shneiderman, B., and Espina, C. (2017). "Classifying Twitter Topic-Networks Using Social Network Analysis," *Social Media+ Society* 3 (1), 2056305117691545.
- Hofman, J. M., Watts, D. J., Athey, S., Garip, F., Griffiths, T. L., Kleinberg, J., Margetts, H., Mullainathan, S., Salganik, M. J., and Vazire, S. (2021). "Integrating Explanation and Prediction in Computational Social Science," *Nature* 595 (7866), 181-188.
- Hu, D., Zhao, J. L., and Cheng, J. (2012). "Reputation Management in an Open Source Developer Social Network: An Empirical Study on Determinants of Positive Evaluations," *Decision Support Systems* 53 (3), 526-533.
- Hunke, F., Engel, C. T., Schüritz, R., and Ebel, P. (2019). "Understanding the Anatomy of Analytics-Based Services—a Taxonomy to Conceptualize the Use of Data and Analytics in Services," *Proceedings of the European Conference on Information Systems*, Stockholm & Uppsala, Sweden.
- Jang, S. M., McKeever, B. W., McKeever, R., and Kim, J. K. (2019). "From Social Media to Mainstream News: The Information Flow of the Vaccine-Autism Controversy in the US, Canada, and the UK," *Health Communication* 34 (1), 110-117.
- Kafeza, E., Makris, C., Rompolas, G., and Al-Obeidat, F. (2021). "Behavioral and Migration Analysis of the Dynamic Customer Relationships on Twitter," *Information Systems Frontiers* 23, 1303-1316.
- Kaiser, J., and Rauchfleisch, A. (2020). "Birds of a Feather Get Recommended Together: Algorithmic Homophily in Youtube's Channel Recommendations in the United States and Germany," *Social Media+ Society* 6 (4), 2056305120969914.
- Kapoor, K. K., Tamilmani, K., Rana, N. P., Patil, P., Dwivedi, Y. K., and Nerur, S. (2018). "Advances in Social Media Research: Past, Present and Future," *Information Systems Frontiers* 20 (3), 531-558.
- Khanam, K. Z., Srivastava, G., and Mago, V. (2023). "The Homophily Principle in Social Network Analysis: A Survey," *Multimedia Tools and Applications* 82, 8811-8854.
- Khurana, I., & Lee, D. J. (2023). "Gender bias in high stakes pitching: an NLP approach," *Small Business Economics* 60, 485-502.
- Kim, Y. A. (2015). "An Enhanced Trust Propagation Approach with Expertise and Homophily-Based Trust Networks," *Knowledge-Based Systems* 82, 20-28.
- Kossinets, G., and Watts, D. J. (2009). "Origins of Homophily in an Evolving Social Network," *American Journal of Sociology* 115 (2), 405-450.

- Kovacs, B., and Kleinbaum, A. M. (2020). "Language-Style Similarity and Social Networks," *Psychological Science* 31 (2), 202-213.
- Kozlowski, A. C., Taddy, M., and Evans, J. A. (2019). "The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings," *American Sociological Review* 84 (5), 905-949.
- Kunst, K., Ringberg, T., and Vatrappu, R. (2022). "Beyond Popularity: A User Perspective on Observable Behaviours in a Digital Platform," *Information Systems Journal* 32 (3), 595-622.
- Kunst, K., and Vatrappu, R. (2019). "Understanding Electronic Word of Behavior: Conceptualization of the Observable Digital Traces of Consumers' Behaviors," *Electronic Markets* 29 (3), 323-336.
- Kwiek, M., and Roszka, W. (2021). "Gender-Based Homophily in Research: A Large-Scale Study of Man-Woman Collaboration," *Journal of Informetrics* 15 (3), 101171.
- Lai, M., Tambuscio, M., Patti, V., Ruffo, G., and Rosso, P. (2019). "Stance Polarity in Political Debates: A Diachronic Perspective of Network Homophily and Conversations on Twitter," *Data & Knowledge Engineering* 124, 101738.
- Laniado, D., Volkovich, Y., Kappler, K., and Kaltenbrunner, A. (2016). "Gender Homophily in Online Dyadic and Triadic Relationships," *EPJ Data Science* 5 (1), 19.
- Lazer, D. M., Pentland, A., Watts, D. J., Aral, S., Athey, S., Contractor, N., Freelon, D., Gonzalez-Bailon, S., King, G., and Margetts, H. (2020). "Computational Social Science: Obstacles and Opportunities," *Science* 369 (6507), 1060-1062.
- Lee, G. M., Qiu, L., and Whinston, A. B. (2016). "A Friend Like Me: Modeling Network Formation in a Location-Based Social Network," *Journal of Management Information Systems* 33 (4), 1008-1033.
- Li, X., Wang, M., and Liang, T. P. (2014). "A Multi-Theoretical Kernel-Based Approach to Social Network-Based Recommendation," *Decision Support Systems* 65, 95-104.
- Li, Y., Gee, W. T. M., Jin, K., and Bond, R. M. (2022). "Examining Homophily, Language Coordination, and Analytical Thinking in Online Conversations About Vaccines on Reddit: A Study Using Deep Neural Network Language Models and Computer-Assisted Conversational Analyses (Preprint)," *Journal of Medical Internet Research*.
- Liu, D., and Brown, B. B. (2014). "Self-Disclosure on Social Networking Sites, Positive Feedback, and Social Capital among Chinese College Students," *Computers in Human Behavior* 38, 213-219.
- Liu, X., Jiang, S., Sun, M., and Chi, X. (2020). "Examining Patterns of Information Exchange and Social Support in a Web-Based Health Community: Exponential Random Graph Models," *Journal of Medical Internet Research* 22 (9), e18062.
- Lynn, T., Rosati, P., Nair, B., and Mac an Bhaird, C. (2020). "An Exploratory Data Analysis of the# Crowdfunding Network on Twitter," *Journal of Open Innovation: Technology, Market, and Complexity* 6 (3), 80.
- Ma, L., Krishnan, R., and Montgomery, A. L. (2015). "Latent Homophily or Social Influence? An Empirical Analysis of Purchase within a Social Network," *Management Science* 61 (2), 454-473.
- Markgraf, M., and Schoch, M. (2019). "Quantification of Echo Chambers: A Methodological Framework Considering Multi-Party Systems," *Proceedings of the 27th European Conference on Information Systems*, Stockholm & Uppsala, Sweden.
- Massachs, J., Monti, C., Morales, G. D. F., and Bonchi, F. (2020). "Roots of Trumpism: Homophily and Social Feedback in Donald Trump Support on Reddit," *Proceedings of the 12th ACM Conference on Web Science*, Southampton, United Kingdom, 49-58.
- McPherson, J. M., and Smith-Lovin, L. (1987). "Homophily in Voluntary Organizations: Status Distance and the Composition of Face-to-Face Groups," *American Sociological Review* 52 (3), 370-379.
- McPherson, M., Smith-Lovin, L., and Cook, J. M. (2001). "Birds of a Feather: Homophily in Social Networks," *Annual Review of Sociology* 27 (1), 415-444.
- Milgram, S. (1967). "The Small World Problem," *Psychology Today* 2 (1), 60-67.
- Mittal, M., Asghar, M. R., and Tripathi, A. (2020). "Do My Emotions Influence What I Share? Analysing the Effects of Emotions on Privacy Leakage in Twitter," *Proceedings of the 19th IEEE International Conference on Trust, Security and Privacy in Computing and Communications*, 1228-1235.
- Möller, F., Stachon, M., Hoffmann, C., Bauhaus, H., and Otto, B. (2020). "Data-Driven Business Models in Logistics: A Taxonomy of Optimization and Visibility Services," *Proceedings of the 53rd Hawaii International Conference on System Sciences*, Waikoloa Village, Hawaii, USA.

- Narang, K., Song, Y., Schwing, A., and Sundaram, H. (2021). "Fuserec: Fusing User and Item Homophily Modeling with Temporal Recommender Systems," *Data Mining and Knowledge Discovery* 35, 837-862.
- Nickerson, R. C., Varshney, U., and Muntermann, J. (2013). "A Method for Taxonomy Development and Its Application in Information Systems," *European Journal of Information Systems* 22 (3), 336-359.
- Noë, N., Whitaker, R. M., Chorley, M. J., and Pollet, T. V. (2016). "Birds of a Feather Locate Together? Foursquare Checkins and Personality Homophily," *Computers in Human Behavior* 58, 343-353.
- Parameswaran, M., and Whinston, A. B. (2007). "Research Issues in Social Computing," *Journal of the Association for Information Systems* 8 (6), 22.
- Qiu, L., Shi, Z., and Whinston, A. B. (2018). "Learning from Your Friends' Check-Ins: An Empirical Study of Location-Based Social Networks," *Information Systems Research* 29 (4), 1044-1061.
- Salganik, M. J. (2019). *Bit by Bit: Social Research in the Digital Age*, Princeton: Princeton University Press.
- Sann, R., and Lai, P.-C. (2020). "Understanding Homophily of Service Failure within the Hotel Guest Cycle: Applying NLP-Aspect-Based Sentiment Analysis to the Hospitality Industry," *International Journal of Hospitality Management* 91, 102678.
- Šćepanović, S., Mishkovski, I., Gonçalves, B., Nguyen, T. H., and Hui, P. (2017). "Semantic Homophily in Online Communication: Evidence from Twitter," *Online Social Networks and Media* 2, 1-18.
- Schaefer, D. R., Kornienko, O., and Fox, A. M. (2011). "Misery Does Not Love Company: Network Selection Mechanisms and Depression Homophily," *American Sociological Review* 76 (5), 764-785.
- Schuetz, S. W., Sykes, T. A., and Venkatesh, V. (2021). "Combating Covid-19 Fake News on Social Media through Fact Checking: Antecedents and Consequences," *European Journal of Information Systems* 30 (4), 376-388.
- Shah, D. V., Cappella, J. N., and Neuman, W. R. (2015). "Big Data, Digital Media, and Computational Social Science: Possibilities and Perils," *The ANNALS of the American Academy of Political and Social Science* 659 (1), 6-13.
- Shalizi, C. R., and Thomas, A. C. (2011). "Homophily and Contagion Are Generically Confounded in Observational Social Network Studies," *Sociological Methods & Research* 40 (2), 211-239.
- Shang, Y., Zhou, B., Zeng, X., Wang, Y., Yu, H., and Zhang, Z. (2022). "Predicting the Popularity of Online Content by Modeling the Social Influence and Homophily Features," *Frontiers in Physics* 10.
- Shmueli, G., and Koppius, O. R. (2011). "Predictive Analytics in Information Systems Research," *MIS Quarterly* 35 (3), 553-572.
- Şimşek, Ö., and Jensen, D. (2008). "Navigating Networks by Using Homophily and Degree," *Proceedings of the National Academy of Sciences* 105 (35), 12758-12762.
- Song, T., Tang, Q., and Huang, J. (2019). "Triadic Closure, Homophily, and Reciprocation: An Empirical Investigation of Social Ties between Content Providers," *Information Systems Research* 30 (3), 912-926.
- Susarla, A., Oh, J.-H., and Tan, Y. (2012). "Social Networks and the Diffusion of User-Generated Content: Evidence from Youtube," *Information Systems Research* 23 (1), 23-41.
- Tang, J., Gao, H., Hu, X., and Liu, H. (2013). "Exploiting Homophily Effect for Trust Prediction," *Proceedings of the 6th ACM International Conference on Web Search and Data Mining*, Rome, Italy, 53-62.
- Tausczik, Y. R., and Pennebaker, J. W. (2010). "The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods," *Journal of Language and Social Psychology* 29 (1), 24-54.
- Todri, V., Adamopoulos, P., and Andrews, M. (2022). "Is Distance Really Dead in the Online World? The Moderating Role of Geographical Distance on the Effectiveness of Electronic Word of Mouth," *Journal of Marketing* 86 (4), 118-140.
- Vaanunu, M., and Avin, C. (2018). "Homophily and Nationality Assortativity among the Most Cited Researchers' Social Network," *Proceedings of 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, Barcelona, Spain, 584-586.
- Vatrapu, R., Mukkamala, R. R., Hussain, A., and Flesch, B. (2016). "Social Set Analysis: A Set Theoretical Approach to Big Data Analytics," *IEEE Access* 4, 2542-2571.

- Wang, Q., Zhao, W., Yang, J., Wu, J., Hu, W., and Xing, Q. (2019a). "Deeprust: A Deep User Model of Homophily Effect for Trust Prediction," *2019 IEEE International Conference on Data Mining*, 618-627.
- Wang, X.-W., Cao, Y.-M., and Park, C. (2019b). "The Relationships among Community Experience, Community Commitment, Brand Attitude, and Purchase Intention in Social Media," *International Journal of Information Management* 49, 475-488.
- Wang, X., and Song, Y. (2020). "Viral Misinformation and Echo Chambers: The Diffusion of Rumors About Genetically Modified Organisms on Social Media," *Internet Research* 30 (5), 1547-1564.
- Waniek, M., and Hidalgo, C. A. (2022). "Bridging the Polarization Gap: Maximizing Diffusion among Dissimilar Communities," *Collective Intelligence* 1 (2), 26339137221128542.
- Watts, D. J., and Strogatz, S. H. (1998). "Collective Dynamics of 'Small-World' networks," *Nature* 393 (6684), 440-442.
- Webster, J., and Watson, R. T. (2002). "Analyzing the Past to Prepare for the Future: Writing a Literature Review," *MIS Quarterly* 26 (2), xiii-xxiii.
- Williams, H. T., McMurray, J. R., Kurz, T., and Lambert, F. H. (2015). "Network Analysis Reveals Open Forums and Echo Chambers in Social Media Discussions of Climate Change," *Global Environmental Change* 32, 126-138.
- Xu, S., and Zhou, A. (2020). "Hashtag Homophily in Twitter Network: Examining a Controversial Cause-Related Marketing Campaign," *Computers in Human Behavior* 102, 87-96.
- Yang, X., Joukova, A., Ayanso, A., and Zihayat, M. (2022). "Social Influence-Based Contrast Language Analysis Framework for Clinical Decision Support Systems," *Decision Support Systems*, 113813.
- Yuan, G., Murukannaiah, P. K., Zhang, Z., and Singh, M. P. (2014). "Exploiting Sentiment Homophily for Link Prediction," *Proceedings of the 8th ACM Conference on Recommender Systems*, Foster City, California, USA, 17-24.
- Zhang, B., Pavlou, P. A., and Krishnan, R. (2018). "On Direct Vs. Indirect Peer Influence in Large Social Networks," *Information Systems Research* 29 (2), 292-314.
- Zhang, J., Wang, W., Xia, F., Lin, Y.-R., and Tong, H. (2020). "Data-Driven Computational Social Science: A Survey," *Big Data Research* 21, 100145.
- Zhang, K., and Pelechrinis, K. (2014). "Understanding Spatial Homophily: The Case of Peer Influence and Social Selection," *Proceedings of the 23rd International Conference on World Wide Web*, Seoul, Korea, 271-282.
- Zhang, X., Zhu, J., Xu, S., and Wan, Y. (2012). "Predicting Customer Churn through Interpersonal Influence," *Knowledge-Based Systems* 28, 97-104.
- Zhou, Z., Xu, K., and Zhao, J. (2018). "Homophily of Music Listening in Online Social Networks of China," *Social Networks* 55, 160-169.