Topic Recipe-based Social Simulation for Research Dynamics Analysis

Keeheon Lee
Department of Information & Industrial Engineering,
Yonsei University, 120-749 Seoul, Korea
Email: keeheon4u@gmail.com

Chang Ouk Kim
Department of Information & Industrial Engineering,
Yonsei University, 120-749 Seoul, Korea
Email: kimco@yonsei.ac.kr

***Abstract*—In this paper, we introduce an agent-based modeling and simulation model for research dynamics analysis. Since researchers constitute research systems in research dynamics, modelling the behavior of a researcher is a key to this method. A researcher makes topic recipes for research products projecting his/her interest and fulfilling financial needs under his/her capability and topical trends. A topic recipe means a combination of topics in a research field. A topic can be related to a methodology or domain knowledge. However, the researcher chooses the favorable topic recipe among the recipes for each. We analyzed the forecasting power of our model. We also examined research dynamics, in terms of foresight, with respect to the social network structure of researchers. In result, we confirmed many topical trends generated by our model had similar pattern to the real topical trends. However, there still are future works to improve the forecasting power.**

Keywords: Social simulation, Research dynamics analysis, Topic recipe, Forecasting, Foresight, Topic modeling

I. Introduction

RESEARCH dynamics is a stream of studies on research that focuses on the evolution of research system over time. It pays attention to the dynamic interactions between constituents of the system. The system consists of researchers with a certain relationship among them. The system yields research products of researchers. The products are scientifically measurable and compositions of topics as in latent Dirichlet allocation [4]. Research products are also input for the individual knowledge function [2] and for the national economic growth function [5]. And thus, the relation between the dynamic changes in the research topics according to the constituents is the interest of research dynamics.

Research dynamics is important to both a researcher and a research policy maker. A researcher could set a competitive strategy against the other researchers [11]. A policy maker would specify operational processes to achieve multiple objectives for national economic growth by utilizing the principles lie in the research dynamics [3].

The vast amount of studies on dynamics analysis has a limit to incorporate a researcher decision-making process along with social interactions in micro level since empirical data limitations exist. But simulation can be the breakthrough for the restriction by uncovering the output of social interactions among researchers deciding what research product to make strategically. We develop researcher agent-based modeling and simulation, which we call social simulation. We believe our simulation for research dynamics analysis not only provides insights for foresight but is capable of capturing topical trends for forecasting.

There are three contributions. First, it extends the present social simulation of computational modeling approach to scientific research dynamics. Second, it suggests social simulation as a foresight tool as well as a forecasting tool for research dynamics analysis. Third, it provides a useful way to alleviate the problem of calibration in social simulation when forecasting topical trends.

II. Literature Review

A. Research Analysis

Scholars have studied scientific research for several objectives: understanding science and technology trends, investigating the relationship among research products and actors in research, scrutinizing the actors' dynamics, assessing the impact of scientific research, forecasting the next research issues, and establishing plans and policies to give incentives to research community to achieve a certain goal of a policy maker .

The scholars utilized various methods to accomplish their goals. The popular methods are estimation by regressing econometric models, input-output analysis, game theory, Delphi, technological road mapping, network analysis, and content analysis with the help of text mining. However, no method has a perfect power to explain phenomena but one method is complementary to another method. For analyzing future, a methodological crossover between fields and the integration of the present methods seems promising [1]. In particular, social simulation as an emerging methodology that simulates complex systems is highlighted.

B. Social Simulation

Social simulation has grown its share throughout disciplines. However, a small number of social simulations are available for research dynamics analysis [9]. Pajares, Lopez, and Hemandez (2003) built a social simulation for industrial sector with an objective to test and to assess firms' strategies, in terms of R&D management [7]. Firms and consumers comprise the artificial society of this social simulation. Firms decide process investment and product innovation strategically. Subsequently, Pajares, Hemandez-Iglesias, and Lopez-Paredes (2004) proposed advanced industrial research dynamics where firms learn decisions on

R&D budget, production, and technology [8]. Cerulli (2012) modeled a stochastic game between a research foundation and a funded firm for an examination of R&D subsidies [6]. The author concluded that a collaborative strategy outweighs a rivalry strategy between the foundation and the firm to reach the social optimum within R&D funding.

Similar social simulations for analyzing various dynamics in a system are on market dynamics, epidemics, emergency dynamics, opinion dynamics, and ecological dynamics.

III. Model

We design a research topic recipe-based decision-making model, consisting of acknowledging the social values for topics, creating the candidate recipes, and choosing one research topic recipe for each time. We build a researcher social simulation model by applying this decision-making model to researcher agents within a social network.

A. Research Topic Recipe Model

Assume a research product is a composition of topics. A researcher considers how to blend topics into a research product. That is, a research product is an implementation of a topic recipe. The researcher would think several candidates of topic recipes and choose the most satisfactory and probable one from the candidates. The researcher may not only refer to individual interests on topics but also consult social values of topics. Social and psychological aspect of getting reputation matters to manufacturing a research product. Self-interest of solving problems also motivates researchers to conduct on making a research product. Money is also a motivational factor to researchers to fabricate a research product. This researcher topic recipe model is divided by four subsequent phases. (1) Acknowledgement of social values of topics: the model starts with assessing social values of topics. (2) Creation of research topic recipes: making research topic recipes is to derive candidate research topic recipes from the topic recipe of the latest research product. (3) Evaluation of the research topic recipes: an agent evaluates the candidates to compare one to another. (4) Selection of one research topic recipe: a researcher agent is likely to produce a research product that would have more valuable.

IV. Experiment and result

We implement our model on the top of NetLogo. We assume the researcher social network is a small world network [12]. We change the rewiring probability to see how social values and topic recipes change according to different social structures. We compare topic modeling result of the abstracts of IIE Transactions with our result by dynamic time warping (DTW). DTW returns the similarity between two different time series data [10].

The first result shows the similarity between the simulation result and topic modeling result. Many are similar in terms of pattern. It implies that the topical trends are the result of an emergent and collective behavior of researchers in research systems. The second result shows the changes in ten topics with respect to the rewiring probability of a social network. The increase in the rewiring probability implies the average path lengths among researchers are shortened and the degree of clustering in researchers becomes higher. It can be done by the development of media such as web.

V. Conclusion

The result of our social simulation shows the potential of its power to forecast and foresee the future topical trends. However, to improve its precision, we may have to conduct more experiments on other variables.

VI. References

[1] Technology future analysis, toward integration of the field and new methods, *Technological Forecasting and Social Changes*, vol. 71, pp. 287-303, 2004.
[2] Z.J. Acs, L. Anselin, A. Varga, “Patents and innovation counts as measures of regional production of new knowledge”, *Research Policy*, vol. 31, pp.1069-1085, 2002.
[3] L.M. Beier, “Incentive, reward, development, or welfare? Revision of an integral grant program”, Journal of Research Administration, vol. 33, no. 1, pp.5-12, 2002.
[4] D. Blei, A. Ng, M. Jordan, “Latent Dirichlet allocation”, *Journal of Machine Learning Research*, vol. 3, 2003.
[5] L. Bornmann, “What is societal impact of research and how can it be addressed? A literature survey”, *Journal of the Association for Information Science and Technology*, vol. 64, issue 2, pp. 217-233, 2013.
[6] G. Cerulli, “Are R&D subsidies provided optimally? Evidence from a simulated agency-firm stochastic dynamic game”, *Journal of Artificial Societies and Social Simulation*, vol. 15, no. 1, pp.7, 2012..
[7] J. Pajares, A. Lopez, C. Hernandez, “Industry as an organisation of agents: Innovation and R&D management”, *Journal of Artificial Societies and Social Simulation*, vol. 6, no. 2, pp.7, 2003.
[8] J. Pajares, C. Hernandez-Iglesias, A. Lopez-Paredes, “Modeling learning and R&D in innovative environments: a Cognitive multi-agent approach”, *Journal of Artificial Societies and Social Simulation*, vol. 7, no. 2, pp.7, 2004.
[9] S. Parinov, C. Neylon, “Science as a social system and virtual research environment”, *Journal of Artificial Societies and Social Simulation*, vol. 14, no. 4, pp.10, 2011.
[10] H. Sakoe, S. Chiba, “Dynamic programming algorithm optimization for spoken word recognition”, *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 26, no.1, pp. 43-49, 1978.
[11] S.M. Stigler, “*Competition and the research universities*”, Daedalus, 122(4), pp.157, 1993.
[12] D.J. Watts, S.H. Strogatz, “Collective dynamics of ‘small-world’ networks”, *Nature*, vol. 393, pp.440-442, 1998.

Miguel, Amblard, Barceló & Madella (eds.) Advances in Computational Social Science and Social Simulation
Barcelona: Autònoma University of Barcelona, 2014, DDD repository <http://ddd.uab.cat/record/125597>

YONSEI UNIVERSITY

Topic Recipe-based Social Simulation for Research Dynamics Analysis

Keeheon Lee
Information and Industrial Engineering
Yonsei University
keeheon4u@gmail.com

Chang Ouk Kim
Information and Industrial Engineering
Yonsei University
kimco@yonsei.ac.kr

ACKNOWLEDGEMENT: This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (2012R1A1A2046061)

INTRODUCTION

Research dynamics is the evolution of research systems over time due to dynamic interactions between constituents of the systems. Understanding research dynamics is important for establishing Science & Technology policy as research is considered as a driver of national growth.

By taking advantage of computational modeling and simulation, we introduce a social simulation model for research dynamics analysis. In the model, agents make topic recipe-based decisions. It is to complement and advance the present method for research dynamics analysis.

We conduct an experiment to verify the forecasting and the foreseeing power of our model. Our result shows that similar patterns to the real patterns were generated by feeding the first topic proportion only. However, the numbers were not precise. From the forecasting, we could foresee how various outcome would be made by controlling a variable of social network. We suspect delicate but abundant trials are required.

This study contributes in three ways:

Extends the present social simulation on research dynamics.
Confirms social simulation as tools for foresight and forecast.
Alleviates the problem of calibration problem in social simulation.

Topic recipe

Formula of topics as ingredients

A researcher i's topic recipe with n topics at time t is represented as a vector, $\boldsymbol{w}_i(t)$

$\boldsymbol{w}_i(t) = (w_{i1}(t), \ldots, w_{in}(t))$

Social values

Popularity of topics

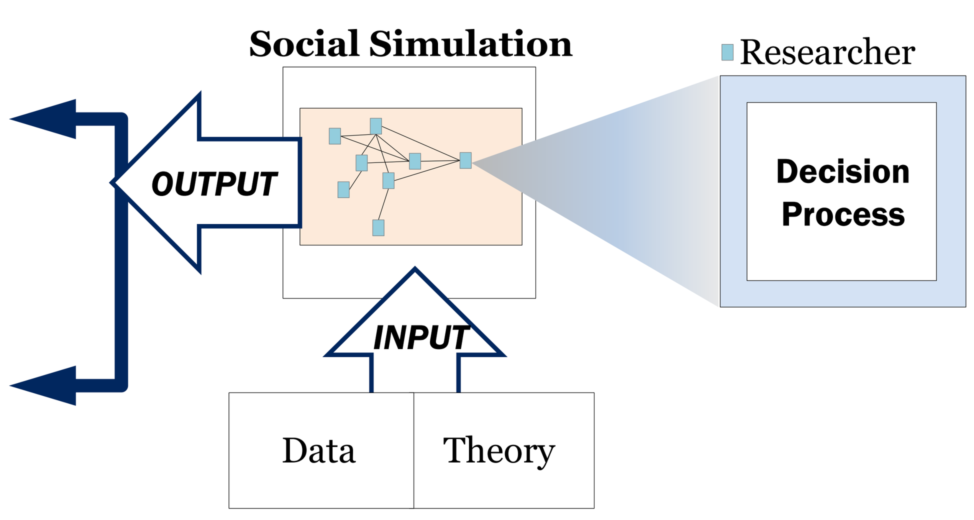
The social values that a researcher i perceived at time t is represented as a vector, $\boldsymbol{x}_i(t)$

$\boldsymbol{x}_i(t) = (x_{i1}(t), \ldots, x_{in}(t))$

MODEL

Forecasting
"What will be"

Foresight
"What might be"



How to "cook" a paper?

(1) Acknowledge what people like
(2) Consider possible recipes
(3) Evaluate the recipes
(4) Choose a favorable recipe
(5) Generate a research product

* We treat producing a research product with topics as cooking food with a recipe.

Description

Agents in our social simulation follow topic recipe-based decision-making process as well as constitute a small world network.

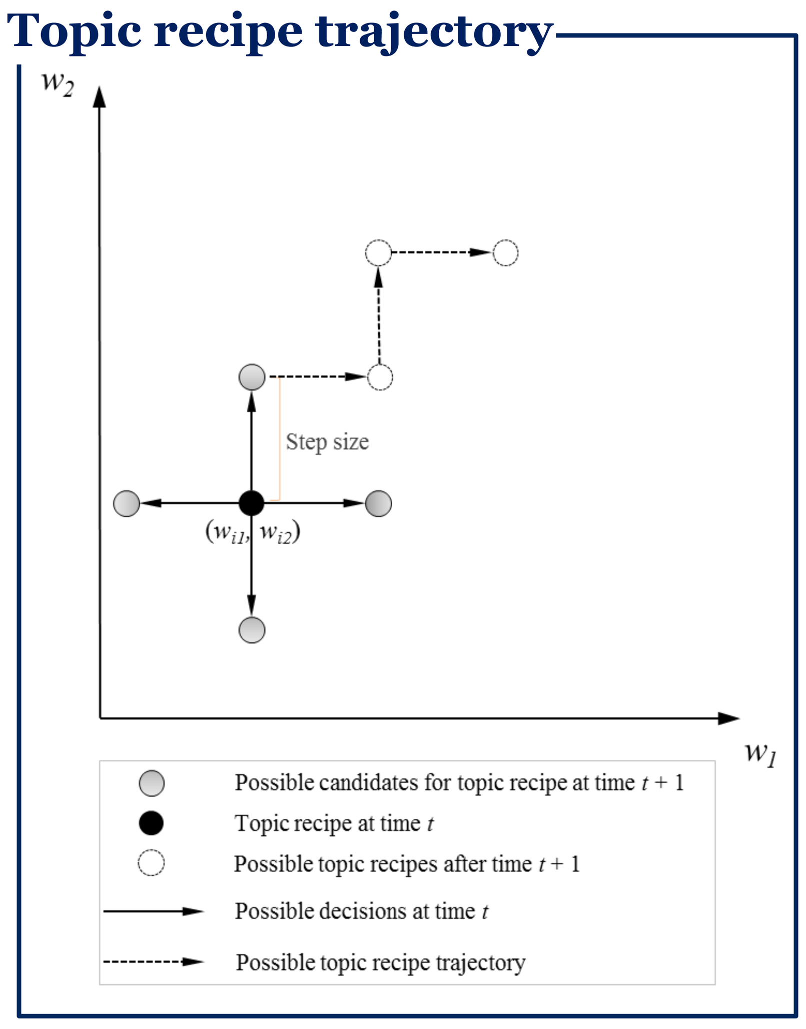
For each time, agents share their social values (Equation 1 and 2) through the network. The values are controlled by social sensitivity, α_i, and social power, of neighbor g, $s_{gi}(t)$.

The agents create possible research topic recipes for the next research product. A topic recipe is equivalent to a combination of weights on topics. Scholarly self-interest and financial interest are the innate factors for generating a new research product (Equation 3). The possible research topic recipes originate from the topic recipe of the current product by random walk approach as in the figure of "Topic recipe trajectory".

And then, the agents evaluate the possible research topic recipes. The scholarly value of a candidate is the weighted sum of the current social values and the scholarly part of every recipe (Equation 4; $p_i^s(t;k)$ is agent i's scholarly value of k^{th} recipe at time t and $w_{ij}^s(t;k)$ is agent i's scholarly weight on j^{th} topic at time t). The financial value of a candidate k is the weighted sum of the fund weight(proportion) and the financial part of every recipe (Equation 5). And, the total value for the candidate k is the weighted sum of them (Equation 6; ω_i^s and ω_i^f are agent i's weight on scholarly and financial values).

Finally, each agent chooses a topic recipe for the next research product probabilistically. The probability of choosing k^{th} candidate at time t is calculated by Multinomial logit rule.

Topic recipe trajectory



LITERATURE REVIEW

RESEARCH DYNAMICS ANALYSIS on

Understanding science and technology trends (Yan, 2014)
Investigating the relationship between research product and researchers (Olmos-Penuela *et al.*, 2014)
Scrutinising actors' dynamics (Martinson *et al.*, 2009)
Assessing the impact of science research (Hall and Reenen, 2000; Lee *et al.*, 2012)
Forecasting the next research issues (Halal, 2010)
Establishing plans and policies (Martin, 2010)

RESEARCH DYNAMICS ANALYSIS by

Econometric models (Kleinknecht and Reijne, 1991)
Input-Output Analysis (Hall and Reenen, 2000)
Game Theory (Dasgupta and Maskin, 1987)
Delphi (Harold and Turoff, 2011)
Technological Road Map (Carvalho *et al.*, 2013)
Network Analysis (Lee *et al.*, 2012)
Text Mining (Blei *et al*, 2003, Yan, 2014)

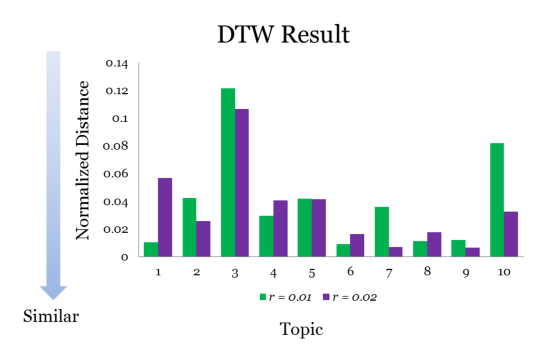
SOCIAL SIMULATION

Industrial sector to test and to assess firms' strategies on investment (Pajares *et al.*, 2003)
R&D budget, production and technology (Pajares *et al.*, 2004)
Stochastic game between research foundation and funded firm for examining R&D subsidies (Cerulli, 2012)

Similar dynamics: Market dynamics (Zenobia *et al.*, 2009; Lee *et al.*, 2013, 2014), epidemics(Burke *et al.*, 2006; Rahmandad and Sterman, 2008), emergency dynamics(Mysore *et al.*, 2006; Narzisi *et al.*, 2006; Carley et al., 2006), opinion dynamics (Deffauant *et al.*, 2001), ecological dynamics (McLane *et al.*, 2011)

RESULT

Forecasting

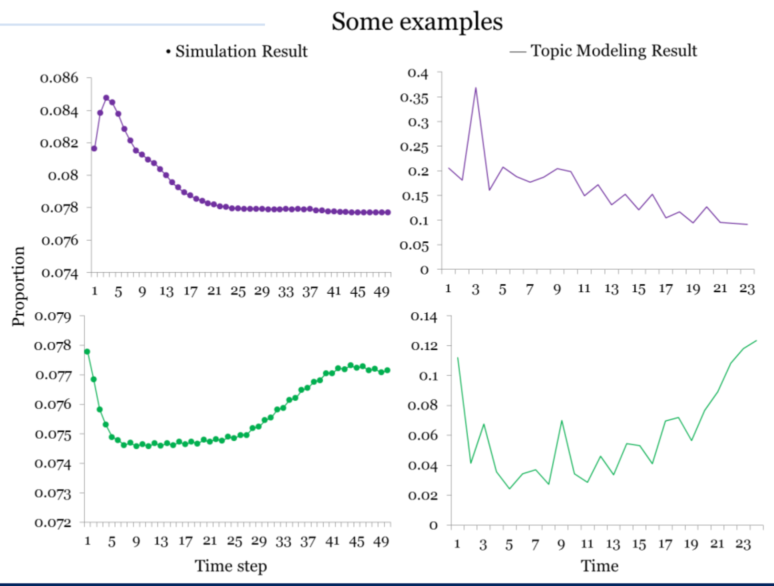


As shown on the left chart, not all but many topical trends were similar between our simulation result and topic modeling result. Scales were different in some degrees but patterns were similar. Our result revealed the fact that research dynamics can be simulated if a proper decision-making process of an agent and a decent social network structure are given. The below figures are two examples of similar patterns (Topic 10 and topic 1).

Data: IIE Transactions (1990~2013)
#Agents: 100
#Topics: 10
α: 0.3
Rewiring probability (r): 0.01~0.5
#Replications: 30
Tool: NetLogo 5.0.5
Topic modeling: LDA (Blei *et al.*, 2003)

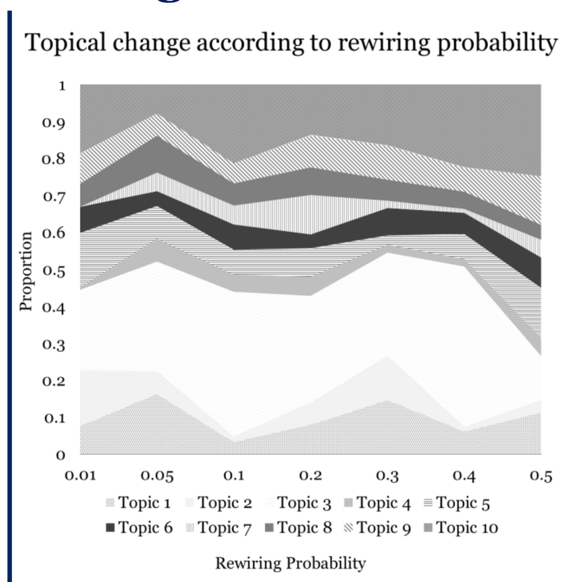
*Initial proportions of 10 topics are given by the result of LDA.

Some examples



Foresight

Topical change according to rewiring probability



The figure on the left shows the changes in the final proportions of ten topics as scholars get clustered. It was done by controlling the rewiring probability of the social network of agents. The rewiring probability shortens average path lengths between agents and clusters agents highly.

The development of media such as web makes the rewiring probability high. And thus, the result depicts the topical changes that might be made by the clustering of researchers.

Topic 1: System-Reliability-Maintenance-Failure-Repair-Inspection-Degradation-Replacement
Topic 2: Method-Data-Model-Simulation-Performance-Design-Analysis-Parameter
Topic 3: System-Performance-Throughput-Machine-Production-Buffer-Line-Flow
Topic 4: Problem-Algorithm-Solution-Programming-Optimal-Heuristic-Computational-Optimization
Topic 5: Production-Cost-Optimal-Inventory-Policy-Demand-Product-Order
Topic 6: Scheduling-Time-Job-Machine-Processing-Batch-Algorithm-Heuristic
Topic 7: Surface-Part-Machining-IIE-Transactions-Edition-Article-Materials
Topic 8: Supply-Demand-Price-Information-Chain-Supplier-Manufactuerer-Retailer
Topic 9: Control-Process-Chart-Quality-Sampling-Performance-Monitoring-Statistical
Topic 10: Manufacturing-System-Design-Model-Process-Assembly-Quality-Decision-Control

Dynamic Time Warping (DTW)

A similarity measure for different size of two time series. The larger value implies the larger difference.

Equations

$$\boldsymbol{x}_i(t+1) = (1-\alpha_i)\boldsymbol{z}_i(t) + \alpha_i \sum_{g \in G_i} s_{gi}(t)\boldsymbol{z}_g(t) \quad (1)$$

α_i: the sensitivity to social influence
g: one in agent i's neighbor set G_i
$s_{gi}(t)$: the degree of social power of g to i at time t
$\boldsymbol{z}_g(t)$: a normalized vector for g's social values at time t

$$z_{ij}(t) = x_{ij}(t)w_{ij}(t) / \sum_{j=1} x_{ij}(t)w_{ij}(t) \quad (2)$$

$z_{ij}(t)$: j^{th} topic value of $\boldsymbol{z}_i(t)$
$x_{ij}(t)$: j^{th} topic value of $\boldsymbol{x}_i(t)$
$w_{ij}(t)$: j^{th} topic value of $\boldsymbol{w}_i(t)$

$$w_{ij}(t) = w_{ij}^s(t) + w_{ij}^f(t) \quad (3)$$

$w_{ij}^s(t)$: j^{th} topic value of the self-interest(scholarly) part of $\boldsymbol{w}_i(t)$
$w_{ij}^f(t)$: j^{th} topic value of the financial interest part of $\boldsymbol{w}_i(t)$

$$p_i^s(t;k) = \sum_{j=1} x_{ij}(t)w_{ij}^s(t;k) \quad (4)$$

$p_i^s(t;k)$: i's self-interest value of k^{th} candidate topic recipe at time t
$w_{ij}^s(t;k)$: i's self-interest weight on k^{th} candidate topic recipe at time t

$$p_i^f(t;k) = \sum_{j=1} d_j(t)w_{ij}^f(t;k) \quad (5)$$

$p_i^f(t;k)$: i's financial interest value of k^{th} candidate topic recipe at time t
$w_{ij}^f(t;k)$: i's financial interest weight on k^{th} candidate topic recipe at time t
$w_{ij}^f(t:k)$: j^{th} topic value of the financial interest part of $\boldsymbol{w}_i(t;k)$
$d_j(t)$: the fund proportion of j^{th} topic

$$p_i(t;k) = \omega_i^s p_i^s(t;k) + \omega_i^f p_i^f(t;k) \quad (6)$$

$p_i(t;k)$: i's total value of k^{th} candidate topic recipe at time t
ω_i^s: i's weight on self-interest value
ω_i^f: i's weight on financial value

$$\Pr\{\boldsymbol{w}_i(t+1) = \boldsymbol{w}_i(t;k)\} = \frac{\exp(p_i(t;k))}{\sum_k \exp(p_i(t;k))} \quad (7)$$

CONCLUSION

Implications

Forecasting – "What will be"

Our social simulation can induce the similar topical trend patterns to the real topical trends with only the very first data of topical proportions. And, it means topical patterns are from an emergent and collective behavior of researchers in research systems. The detail fluctuation, which can be called noise for now, is not presented in simulation result. It means there still is a room for advancement in the power to forecast topical trends. This advancement may empower the precision of the numbers for topical proportions for each time.

Foresight – "What might be"

Our social simulation can provide how the output may change according to the structural modification in the social network of researchers. This may help a policy maker to decide the extent of coordination between researchers.

Future Work

Limitation

Limited control variables were applied: rewiring probability was the only consideration.

Social power of each agent was identical and constant: $s_{gi}(t)$ was set by the inverse of the number of neighbors and was invariant.

Assume topic model generates good quality of topic assignments.

Future research topics

Not only rewiring probability but also the sensitivity to social influence, the proportion of self-interested part of a topic recipe, the weights of self-interest and financial interest values, the stride of random walk, the number of candidates, the number of agents, the number of topics, and the rule for social influence can be controlled.

Several different topic modelling results can be used as benchmarks.

Reference

Technology future analysis, toward integration of the field and new methods, *Technological Forecasting and Social Changes*, vol. 71, pp. 287-303, 2004.
K.M. Carley, "Computational modeling for reasoning about the social behavior of humans", *Computational and Mathematical Organization Theory*, vol. 15, pp. 47-59, 2009.
G. Cerulli, "Are R&D subsidies provided optimally? Evidence from a simulated agency-firm stochastic dynamic game", *Journal of Artificial Societies and Social Simulation*, vol. 15, no. 1, pp.7, 2012.
D. Blei, A. Ng, M. Jordan, "Latent Dirichlet allocation", *Journal of Machine Learning Research*, vol. 3, 2003.
J. R. Harrison, Z. Lin, G. R. Carroll, K.M. Carley, "Simulating modeling in organizational and management research", *The Academy of Management Review*, vol. 32, No. 4, pp. 1229-1245, 2007.
A. Lam, "What motivates academic scientists to engage in research commercialization: 'Gold', 'ribbon', or 'puzzle'?", *Research Policy*, vol. 40, pp.1354-1368, 2011.
J. Pajares, A. Lopez, C. Hernandez, "Industry as an organisation of agents: Innovation and R&D management", *Journal of Artificial Societies and Social Simulation*, vol. 6, no. 2, pp.7, 2003.
J. Pajares, C. Hernandez-Iglesias, A. Lopez-Paredes, "Modeling learning and R&D in innovative environments: a Cognitive multi-agent approach", *Journal of Artificial Societies and Social Simulation*, vol. 7, no. 2, pp.7, 2004.
S. Parinov, C. Neylon, "Science as a social system and virtual research environment", *Journal of Artificial Societies and Social Simulation*, vol. 14, no. 4, pp.10, 2011
E.R. Smith, F. R. Conrey, "Agent-based modeling: A new approach for theory building in social psychology", *Personality and Social Psychology Review*, vol.11, no.1, pp.87-104, 2007.