Strategic Foundation of Computational Social Science

Yunkyu Sohn*
Supervised by: James Fowler and Sue Moon
Korea Advanced Institute of Science and Technology
University of California San Diego
La Jolla, California 92093, USA
ysohn@ucsd.edu

ABSTRACT

For decades, scholars of various disciplines have been fretted over strategic interactions, presenting theoretical insights and empirical observations [3, 18, 25]. Despite the central role played by strategic interactions in creating values in the Internet environment, our ability to understand them scientifically and to manage them in practice has remained limited. While engineering communities suffer from not having enough theoretical resource to formalize such phenomena, economics and social sciences lack adequate technology to properly operationalize their theoretical insights, thereby demanding an integrative solution. This project aims to develop a rational-choice-theory-driven framework for computational social science, focusing on social interactions on the Internet. In order to suggest theoretical foundations, validation of the predictions in a controlled environment, and verification of the results in actual platforms, general approaches and a few examples of ongoing research are presented.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous; J.4 [**Social and Behavioral Sciences**]

Keywords

rational choice theory, social networks, online behavior

1. PROBLEM

Incentivizing individuals to release information and stimulating their spontaneous interactions are at the core of managing human-aided computing platforms (e.g.information filtering systems) and value sharing on social networking services. At the moment, however, we lack a common device, which enables precise description of their underlying dynamics.

*Reseach scientist at KAIST and PhD student at UCSD.

WWW'14 Companion, April 7–11, 2014, Seoul, Korea.
ACM 978-1-4503-2745-9/14/04.
http://dx.doi.org/10.1145/2567948.2567962.

In order to approach this issue, we need to develop a research program which provides an integrative solution to understand and manage human interactions. Foremost, we need to devise a theoretical foundation which provides language for analyzing user interactions. While many studies in computational social science incorporate an extant of indices developed in social sciences, the relationship between actual human behavior and the indices are unclear. Hence, in addition to providing a framework for understanding online social interactions, theoretical efforts are expected to reveal what it means to use such indices in different classes of platforms and interactions. Second, the theory-driven language must concord with our knowledge on realistic behavior of human. Since online social behavior also have to be explained by the general theory of social action, empirical observations of social sciences should be seriously considered in computational social science research. In particular, experimental studies in controlled environments enable us to investigate explicit roles of treatment variables that are very difficult to identify with observational data. On the basis of experimental resources, we can approximate extremely complicated real phenomena by separating key variables and their relationships. In order to analyze such data, technical tools developed in statistics and engineering communities must be used, particularly for working with massive datasets. Finally, all this combined information should lead us to suggest practical implications for managing the highly interconnected systems.

2. STATE OF THE ART

Theory Rational choice theory is a foundational approach in behavioral science as it theorizes motivations behind behavior with parsimonious model-based frameworks. In most of the models, researchers assume self-interested individuals who aim to maximize their utilities by conducting strategic calculus. Theoretical models relevant to computational social science are well-summarized in [9] and [12]. Despite such conceptual advances, the framework has been merely applied to analyzing very few scenarios of online phenomena. Rare examples include evolution of cooperation in P2P systems [26, 11], reputation effects in online auctions [8] and resource allocation problems in computer networks [16].

Experiment Theory-driven predictions may deviate from actual human behavior. In order to ensure relevance of theoretical predictions as empirical models of behavior, we need to study actual human behavior in a controlled setting. Most experimental studies relevant to the present project are from psychology and economics. In the course of conducting laboratory experiments, social scientists realized the difference between theoretical predictions and actual human behavior. Famous examples include loss aversion and inter-temporal discounts. Without an exception, as a system of cognitively biased individuals who do not follow mathematical simplicity, experimental results from traditional disciplines must be seriously incorporated when explaining online social datasets. A recently introduced promising approach is using the Internet as a platform for recruiting human subjects and conducting experiments. Several studies incorporate randomized experimental techniques to gauge the extent of peer effects on social networks [2], and some other studies use online labor markets to observe the patterns of strategic interactions among experimental subjects [22].

Measurement and Observation Many studies in computer science community aims to measure and observe social interactions on the Internet. In order to properly quantify what we want to know, researchers have developed analytic techniques for characterizing structural and dynamic properties of the phenomena and extracting information from them, suited for massive web data analysis. In particular, the empirical estimation techniques incorporate data mining, natural language processing, machine learning and graph analysis algorithms. By utilizing such techniques many interesting observations are reported [15]. However, these studies often had unclear connections with theoretical insights and analytic models of human behavior, suggesting anecdotal evidence rather than theory-driven explanations of human behavior. As a consequence, systemic understanding of online social interactions has been difficult.

3. PROPOSED APPROACH

The core value of the present approach is in that we assemble theoretical insights and empirical evidence to promote practical management of social systems. First, I propose theoretical models, which simultaneously incorporate rational and bounded-rational models of mind. Second, such assumptions and predictions are tested in controlled experimental environments. Third, Internet field experiments and observational studies are conducted in order to obtain external validity of the results derived from the experiments. Finally, I suggest that the integrated results can be used to design incentive structure for nudging users to desired directions and provide welfare relevant criteria for platform design.

4. METHODOLOGY

While heavily relied on sociological theories of action, existing measures of social interactions lack autonomous explanation on why someone does what. For example, despite the fact that peer effect has become the most common subject of research in social network analysis on massive web data, motivations and utility behind the observed statistics remain vague. Moreover, while algorithmic approaches for influence maximization exist [17], we are uncertain how a commonly used node centrality measure represents the importance of a person or a relationship with respect to influence in a given context.

The connections between individual behavior, collective outcomes, and unobserved decision process must base upon rigorous and parsimonious theories of human behavior. Among possible forms of such theories, the most convincing and useful approach for quantitative research is mathematical modeling.

Predictions of theoretical models are tested with behavioral experiments in controlled environments. Computer and Internet platforms are actively incorporated. Obtained results are expected to provide information that is not entailed in the theoretical studies.

The proposed approach aims to introduce theoretical models in connection with experimental and observational studies on the Internet. In addition to the traditional economic modeling of human behavior, as psychological reasoning of human agents do not guide to ideal choices, theories of learning and refined equilibrium concepts are actively incorporated from evolutionary game theory and theoretical ecology.

5. RESULTS

At the current stage of the project, several theoretical and experimental studies are conducted. The ultimate goal is to utilize the present understanding of human behavior to explain and predict massive social interactions on the Internet, and to propose suggestions for future Internet development.

In this note, I highlight two classes of network node centrality measures in a strategic perspective. First, I show that individual best response action in a particular type of simultaneous move game corresponds to a general form of recursive centrality measure. Second, I present results from a laboratory experiment showing the impact of partner selection on evolution of cooperation and how this dynamics explains k-core differentiation of nodes in temporally evolving networks. Additionally, I propose a bipartite network formation model which aims to explain mutualistic cooperation between services and platforms, or users and services.

5.1 Game Theory: Strategic Foundation of Recursive Node Centrality

An extant number of research utilize centrality measures as their key variables to explain structural substrates of network dynamics. However many studies fail to present how their theoretical constructs relate such quantitative indices.

Among popular indices of social network analysis, I focus on recursive node centrality measures which include the famous Bonacich, Katz, and PageRank centrality. As widely recognized, these measures have been in great success for link predictions and recommendations in web search engines. While initially proposed to gauge the extent of influence one can exert in a network, strategic foundation of such measures has been only recently approached [4].

In order to examine strategic origin of recursive centrality measures, I study a non-cooperative model of induced preference, where external actors aim to influence internal actors while connected internal pairs have conformity bias. Assuming a quadratic utility function in an ideological or preference space, a formal model of external influence under internal conformism consists of an internal member's utility function, $u_i : a_i \times \mathbf{a} \to \mathbb{R}$ which is defined as:

$$u_i = -\pi \sum_{k \in M} B_{ik}(a_i - b_k)^2 - (1-\pi) \sum_{j \in N} A_{ij}(a_i - a_j)^2 \quad (1)$$

where $\frac{\pi}{1-\pi}$ indicates the relative importance of the total amount of external influence over the total amount of internal influence, internal interaction matrix A, external influence matrix B, external actor's preferred position vector, $\mathbf{b}$, and internal actor's induced position vector, $\mathbf{a}$, and an external member's utility function,

$$u_k = -\frac{1}{N}\sum_{i\in N}(a_i - b_k)^2 - c_k, \qquad (2)$$

where $c_k = f(\sum_{i\in N} B_{ik}) \propto \sum_{i\in N} B_{ik}$, indicating that the amount of investment external actor k pays to have influence over internal actors increases as his volume of connections grows up.

I first analyze a simultaneous move n-player game where external actor preference is given and only internal actors make decisions. When individual best response actions (Nash equilibria) are assumed, internal actors' positions follow a typical form of recursive centrality measure, $\mathbf{a}^* = (I - (1-\pi)A)^{-1}\pi B\mathbf{b}$. Notably, this equilibrium deviates from utilitarian welfare maximization equilibria. So called price of anarchy, the deviation of aggregate utility in Nash equilbria from global maximum aggregate utility value, is examined.

I proceed to examine strategic incentives of external actors in a sequential move game, where external actors move first and then internal actors move in response to the determined B matrix, presenting implications for optimal targeting strategy for those wanting to perturb a socially interconnected system. In this game, cost-effective investment strategy of external actors is shown. A striking finding is that, over a large domain of $f(\sum_{i\in N} B_{ik})$, an external actor's utility increases as one's targets' connection profiles become less redundant.

This finding offers an answer to a puzzle in political clientelism why both friendly and counteractive lobbying are prevalent [7]. When organized interests aim to buy a legislature by purchasing leadership, they target a combination of actors, not based upon aggregate individual centrality, but those having the most diverse set of partners.

My model can serve as a basic framework for analyzing massive herding behavior in social networking services. For example, this model can illustrate what happens if a celebrity or a politician announces support for a particular policy on Facebook. Also my model suggests an optimal dispersion strategy for a viral marketer for selecting initial adopters of a product. As the model depicts people's preference on a continuum, this model is naturally connected with multidimensional scaling techniques for choice data.

5.2 Evolutionary Game Theory: Social Learning Model for Mutualistic Network Formation

An important, yet under-investigated form of interactions in the social network community is multi-mode networks, where multiple types of nodes interact with different types. Especially, mutualistic interactions between two sets of species (e.g. animals and plants) have been extensively studied in ecology, as it is related to the problem of biodiversity and ecosystem management [5, 6]. Such characterization was also recently applied to explain the dynamics of man-made organizations [24].

On the Internet, mutualistic interactions are pervasive. From platforms to services and services to users, many interactions can be modeled as interactions between two sets of actors. Commonly, actors in the same set compete with each other while actors of different types help each other. Such interactions create recursive and complicated dynamics of wealth transfer and growth.

While the majority of models of mutualistic networks assume passive dynamics in fixed networks. Our research incorporates evolutionary game theoretic formulation to model partner selection dynamics and mutualistic network formation. The results show that empirically observed high network nestedness (low degree animals interact with high degree plants) and dispersed degree distribution are only obtained when people learn others' behavior on the basis of concave utility subject to payoff.

This result suggests that risk averse social learning plays a fundamental role in producing realistic mutualistic interactions. However, we are unaware of mutualistic network strcuture between users and multiple Internet service platforms. Most previous research studies user interactions in a single service platform. While it is suspected that user-service network would also exhibit nested structure, as less active users are expected to use the most popular services, our model suggests that platform and service migration of users would involve risk averse assessment of newly introduced items.

5.3 Laboratory Experiment: Generative Mechanism of Social Capital Entails k-core Differentiation

Despite its popularity in social science research, social capital has hardly had a clear-cut definition. The diverse set of perspectives, that have been suggested by scholars from various disciplines, has yielded an increasing ambiguity in its concept [21]. As a result, existing research commonly incorporates derivatives that are argued to be consequential products of social capital rather than utilizing a direct measurement theory.

Sohn, Choi and Ahn [27] conduct laboratory experiments, in which human subjects play multi-round pairwise prisoner's dilemma game with partner selection option, to trace the generative mechanism of social capital. We show that a popular metric in graph theory called k-core takes a fundamental role in characterizing the generative mechanism of social capital. Temporally stable high-core associations correspond to enduring communities embedding social capital, and emerging community prosperity is associated with structural evolution of the interaction networks.

This research has important implications to relationship maintenance on the Internet, which can be explained by strategic linking behavior of individuals. As shown by user survey and observational studies of unfollow dynamics in Twitter [13, 14], reciprocity centers the utility of maintaining a relationship. Since the interactions between reciprocity concerning individuals can be modeled by a non-cooperative game, strategic models of relationship maintenance can serve as a foundational approach for understanding empirical dynamics. As the asset value of a social networking service embeds in the number of connections, and well-maintained interconnections breed further recruitment of new users, our finding has a core implication in platform design and community management.

6. CONCLUSIONS AND FUTURE WORK

The fundamental value of the present approach is in dissecting the utility of creating prosocial interactions. Ultimately, this effort is linked to formalizing incentive structure in web platforms and understanding what motivates people to participate and cooperate in collective problem solving [10, 19]. In order to achieve this goal, I develop parsimonious mathematical models suited for different strategic scenarios, and conduct online behavioral experiments to test the models in controlled environments. For example, such endeavor help us know what makes certain communities vibrant. Asking whether rewarding avid participators or punishing uninterested users is efficient for facilitating community prosperity [23].

While not extensively covered in this note, proper measurement and analysis of complex datasets always demand computationally efficient, and at the same time, accurate methodological tools. My project also involves technical development of graph analysis methods. The ongoing research includes suggesting a fast and unbiased aggregation schemes for network autocorrelation models and understanding factors determining commonality and bifurcation of graph clustering algorithms [28].

To conclude, this project proposes a unified framework for computational social science research by viewing strategic interaction as a key component of social phenomena. The current stage of the project requires incorporation of observational evidence for verification of theoretical and experimental results. In particular, specific tasks include: 1) tracing temporally varying community association and analyzing their structural properties; 2) measurement of cooperativeness of an individual user; 3) mapping individual cultural and political preference in a latent ideological space, and quantifying peer effects in dynamic data; 4) constructing mutualistic networks between users and services, and services and platforms.

Finally, in order to create socially just and efficient Internet communities, these theoretical and empirical resources should guide us to suggestions for what is good and fair. The classic problem of diversity and growth is a suitable example. "The rich get richer" is a characteristic phrase capturing a host of phenomena observed throughout the evolution of the Internet industry. Facilitated by platform standardization and coordination motives among users to improve communication efficiency, products and ideas are likely to converge. A fundamental question linked to this phenomenon is the paradox of diversity as it creates both problems and opportunities [20]. Simultaneous presence of social integration and population diversity is a difficult goal to achieve. On the Internet, people are connected, yet they are mostly connected to similar others [1]. Also most people use a very few number of platforms yet, at the same time, new services are introduced every day. Coupled with the problem of ecological diversity, thus with the mutualism study introduced above, our efforts should let us predict consequences of the current form of Internet dynamics including this example, and to what extent Internet policy can shape the future of this complex system.

7. ACKNOWLEDGMENTS

This research was supported by Next-Generation Information Computing Development Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Science, ICT & Future Plannig (NRF-2012M3C4A7033342).

8. REFERENCES

[1] L. A. Adamic and N. Glance. The political blogosphere and the 2004 us election: divided they blog. In *Proceedings of the 3rd international workshop on Link discovery*, pages 36–43. ACM, 2005.

[2] S. Aral, L. Muchnik, and A. Sundararajan. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences*, 106(51):21544–21549, 2009.

[3] R. M. Axelrod. *The complexity of cooperation: Agent-based models of competition and collaboration*. Princeton University Press, 1997.

[4] C. Ballester, A. Calvó-Armengol, and Y. Zenou. Who's who in networks. wanted: the key player. *Econometrica*, 74(5):1403–1417, 2006.

[5] J. Bascompte, P. Jordano, C. J. Melián, and J. M. Olesen. The nested assembly of plant–animal mutualistic networks. *Proceedings of the National Academy of Sciences*, 100(16):9383–9387, 2003.

[6] J. Bascompte, P. Jordano, and J. M. Olesen. Asymmetric coevolutionary networks facilitate biodiversity maintenance. *Science*, 312(5772):431–433, 2006.

[7] A. Bonica. Ideology and interests in the political marketplace. *American Journal of Political Science*, 2013.

[8] C. Dellarocas. The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management Science*, 49(10):1407–1424, 2003.

[9] D. Easley and J. Kleinberg. *Networks, crowds, and markets*. Cambridge Univ Press, 2010.

[10] J. H. Fowler and N. A. Christakis. Cooperative behavior cascades in human social networks. *Proceedings of the National Academy of Sciences*, 107(12):5334–5338, 2010.

[11] P. Golle, K. Leyton-Brown, I. Mironov, and M. Lillibridge. Incentives for sharing in peer-to-peer networks. In *Electronic Commerce*, pages 75–87. Springer, 2001.

[12] M. O. Jackson. *Social and economic networks*. Princeton University Press, 2010.

[13] H. Kwak, H. Chun, and S. Moon. Fragile online relationship: a first look at unfollow dynamics in twitter. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1091–1100. ACM, 2011.

[14] H. Kwak, S. B. Moon, and W. Lee. More of a receiver than a giver: Why do people unfollow in twitter? In *ICWSM*, 2012.

[15] D. Lazer, A. S. Pentland, L. Adamic, S. Aral, A. L. Barabasi, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, et al. Life in the network: the coming age of computational social science. *Science*, 323(5915):721, 2009.

[16] R. T. Ma, D. M. Chiu, J. Lui, V. Misra, and D. Rubenstein. Internet economics: The use of shapley value for isp settlement. *IEEE/ACM Transactions on Networking (TON)*, 18(3):775–787, 2010.

[17] R. Narayanam and Y. Narahari. A shapley value-based approach to discover influential nodes in

social networks. *IEEE Transactions on Automation Science and Engineering*, (99):1–18, 2010.

[18] M. A. Nowak. Five rules for the evolution of cooperation. *Science*, 314(5805):1560–1563, 2006.

[19] E. Ostrom. *Governing the commons: The evolution of institutions for collective action*. Cambridge University Press, 1990.

[20] S. E. Page. *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies (New Edition)*. Princeton University Press, 2008.

[21] R. D. Putnam. *Bowling alone: The collapse and revival of American community*. Simon and Schuster, 2001.

[22] D. G. Rand. The promise of mechanical turk: How online labor markets can help theorists run behavioral experiments. *Journal of Theoretical Biology*, 299:172–179, 2012.

[23] D. G. Rand, A. Dreber, T. Ellingsen, D. Fudenberg, and M. A. Nowak. Positive interactions promote public cooperation. *Science*, 325(5945):1272–1275, 2009.

[24] S. Saavedra, F. Reed-Tsochas, and B. Uzzi. A simple model of bipartite cooperation for ecological and organizational networks. *Nature*, 457(7228):463–466, 2008.

[25] T. C. Schelling. *The strategy of conflict*. Harvard University Press, 1960.

[26] J. Shneidman and D. C. Parkes. Rationality and self-interest in peer to peer networks. In *Peer-to-Peer Systems II*, pages 139–148. Springer, 2003.

[27] Y. Sohn, J.-K. Choi, and T. Ahn. Core-periphery segregation in evolving prisoner's dilemma networks. *arXiv:1105.0515*, 2011.

[28] Y. Sohn, J. Ju, H. Kwak, and S. Moon. Exploring the solution sapce of community detection methods. In *STATPHYS*, 2013.