Big Data Research ••• (••••) •••–•••

Contents lists available at ScienceDirect

Big Data Research

www.elsevier.com/locate/bdr

Big Data for Open Digital Innovation – A Research Roadmap ☆

Sabine Brunswicker [a], Elisa Bertino [b], Sorin Matei [c]

[a] Research Center for Open Digital Innovation and Department of Technology Leadership and Innovation, Purdue University, United States
[b] Cyber Center and Computer Science Department, Purdue University, United States
[c] Brian Lamb School of Communication and Cyber Center, Purdue University, United States

ARTICLE INFO

Article history:
Received 29 August 2014
Accepted 4 January 2015
Available online xxxx

Keywords:
Data cyberinfrastructure
Data linkage
Digital collaboration
Digital innovation ecosystem

ABSTRACT

Digital technologies have fundamentally altered the nature of organizing innovation and production leading to open collaboration ecosystems. Individuals self-organize in open, voluntary technology-enabled collectives to share their enhancements to the data or collaborate on analyzing, disseminating, or leveraging the data for many applications, from enterprise computing to mobile, consumer oriented applications. 'Big data' is an increasingly important 'engine' to better understand the complex 'nervous system' of open collaboration. However, we need to equip open collaboration researchers with new datasets that span different contexts, as well as novel computational models and analytical techniques. In this paper, we will elaborate on research questions concerning open digital collaboration and derive the data analytical challenges that need to be addressed to answer these research questions.

© 2015 Elsevier Inc. All rights reserved.

SUBJECTED PROOF

1. Introduction

Digital technologies have changed the fabric of organizations, triggering novel organizational forms. Innovation and production are not confined to established organizations with clear cut boundaries. Individuals self-organize in open, voluntary technology-enabled collectives to share data and knowledge and to jointly create novel solutions for a bewildering array of applications. With these geographically dispersed groups the idea of open source has moved beyond open source software. Manifold types of these collective forms of innovation and production have emerged in which a large number of actors interact and create goods across multiple platforms, contexts, and timelines. These systems increasingly relate to socially significant domains such as health support or eScience. We refer to them as 'open collaboration' systems for innovation and production [24,7,11] and argue that they are dynamic sociotechnical systems for two reasons: (1) they are fluid and (2) work, organization, and technologies are intertwined within them [32]. Wikipedia, Amazon review systems, science network NanoHub.org, and CancerCare.org are just a few examples of open collaboration. Their emergence stimulated researchers to study their nature.

However, most contributions fail to move beyond existing theories and routine application of research tools to tackle the dynamic and sociotechnical reality of open collaboration [13,30]. We argue that 'big data' is an increasingly important 'engine' to make this turn and to better understand the complex 'nervous system' of open collaboration. The interaction in digital environments creates a gigantic stream of behavioral data that provide novel research opportunities to move beyond outdated theories. To do so, we need to equip researchers of open collaboration with new datasets about dynamic sociotechnical processes that span different contexts and users, with novel analytical techniques, and with an efficient and effective research infrastructure to support the development of novel empirically grounded theories and predictive models.

1.1. Open collaboration – an emerging research area of high significance

Open collaboration (OC) relies on a large number of goal-oriented yet loosely coordinated participants, who interact to create a product (or service) of economic value, which is made available to contributors and non-contributors alike. Indeed, OC will have significant economic as well as social impact [24,2,20]. Scholars in social and behavioral science and economics have started to address the emerging phenomena from different angles and present relevant empirical insights. For example, research drawing upon network theory tackles structural characteristics of large OC networks and also presents new insights in the role of authority, reputation, and trust in OC [29]. Contributions in organizational studies and information systems have revealed deeper insights into the factors that motivate individuals to contribute, and the social value they might have [13].

☆ This article belongs to Visionds on Big Data.

E-mail addresses: sbrunswi@purdue.edu (S. Brunswicker), bertino@cs.purdue.edu (E. Bertino), smatei@purdue.edu (S. Matei).

http://dx.doi.org/10.1016/j.bdr.2015.01.008
2214-5796/© 2015 Elsevier Inc. All rights reserved.

1.2. Research gaps and challenges

While this work on OC makes important contributions, major significant gaps remain in understanding, explaining, and predicting the dynamic and sociotechnical nature of OC. Our own research and results from an expert workshops with a community of social science researchers performed in the KredibleNet project [30, 27] suggest three research major areas that call for a multidisciplinary research approach to move beyond a few outdated theories, a snapshot of data without context, and routine tools:

1. *Capturing the multilevel, multimodal and the dynamic nature of roles in OC through multi-modal and pattern data.* OC often implies the interaction of a very diverse set of actors embedded in different contexts, technological platforms and subpopulations. OC is a multimodal and pluri-actor process connecting people through the products and services they create [3,10,32]. Such multi-modal networks form around ties between individuals, their resources, and multiple goods, rather than social relationships [30]. The phenomenon of OC is not confined to one context as individuals are embedded in different contexts and subgroups. Thus, it is crucial to consider linked datasets covering different perspectives and communities hosted on different technological platforms or connected through different goods. In addition, roles are dynamic, fluid, and often enacted in the moment. Thus, a simple structural perspective is not enough. Novel analytical techniques that capture dynamics and consider the interrelation between individual behavior, goods, and group activities offers create opportunities to tackle this gap and understand the inner working of OC.
2. *Acknowledging and unwrapping the constituting role of technology in OC through novel data and analytics.* Researchers regularly 'blackbox' technology even though it holds a constituting role [6,28]. A deeper understanding is thus required of how technological features act as "shapers" of behavior, and enable them to solve complex social problems or create novel services. Indeed, existing OC environments make use of sophisticated features such as recommendation systems (e.g. Amazon review), or visualization techniques (e.g. tag clouds, network visualization) but little is known about how different descriptive or predictive features shape collaboration. Neither do we have sufficient data and analysis about user and technology interaction nor do we sufficiently understand their effect. We require novel data-driven research that captures micro-level data about technologies in different contexts and environments in order to develop novel theories, explanations of collaboration and innovation in OC.
3. *Unpacking the dynamic drivers of performance and sustainability of OC through novel computational models and analytics.* OC has significant implications on established assumptions about organizational forms and their performance implications for firms, society, nations, and beyond. With the increasingly digital economy, the open and collaborative models become economically more viable [2]. At the same time, there is insufficient understanding on what ensures sustainability of OC and the dynamics that drive innovation and performance within it. The availability of longitudinal data and novel computational models that unpack the complex dynamic processes within OC would provide novel insights about how such systems evolve and sustain themselves. Today, we lack a deep understanding of the dynamics at multiple levels.

It is critical to address these challenges and opportunities. However, we need to encourage researchers to take turn and move beyond a few theories, a snapshot of data, and standard research tools. Thus, we need new datasets about dynamic sociotechnical behaviors and novel tools and an efficient and effective research infrastructure to break new grounds in explaining and predicting OC and its impact on innovation.

1.3. Data, infrastructure and novel computational models and analytical tools

Addressing the gaps and data opportunities mentioned above – and the manifold research questions that may arise within them – requires novel datasets on OC embedded in a collaborative infrastructure, and a portfolio of research tools and computational models to analyze this data. There are three major building blocks:

1. *Datasets*: There are today many datasets that can be leveraged for research in the area of OC, including existing and processed data on Wikipedia, open source software development platforms, OC systems forming around platforms like OpenData.gov, as well as data on virtual science infrastructures like NanoHub.org, and open innovation networks like Ninesights. It is crucial however that new datasets be made available, possibly through derivation from existing datasets by mining, experiments, novel processing, and existing and new datasets be linked to create context-rich datasets.
2. *Computational and analytical tools*: There is today a large number of such tools, including metanetwork models, network discovery, dynamic and predictive statistical network analysis, genetic computation, network analysis algorithms, agent-based simulations, sequencing analysis and statistical prediction, event study tools, and collaboration and visualization tools. It is critical however that such tools be easily integrated and made available on unified digital platforms.
3. *Collaborative cyberinfrastructure*: Their goal is to serve as a virtual living lab for experiments, offering the community to build capacity, to share data and results, and communicate findings seamlessly across different media.

1.4. Paper organization

In what follows we first discuss the major research themes in OC. For each such theme we identify guiding research questions and data analytic challenges. We then discuss the relevant features of toolkits needed to address such data analytic challenges. We will then outline a few concluding remarks.

2. Research themes in open collaboration

Fig. 1 presents our framework that proposes a dynamic sociotechnical system perspective to a roadmap towards a multidisciplinary research on OC and the identification of data analytic challenges. Our framework conceptualizes OC systems as a sociotechnical system which subsume loosely individuals which freely contribute to develop novel goods, with particular resources and technological features [24]. It addresses three major research dimensions, namely (1) multi-modal and dynamic roles, (2) technological affordances, and (3) the performance and sustainability of the overall dynamic system (as well as the actors and goods within them). These three dimensions will guide our research roadmap.

2.1. Thematic area 1: multilevel, multi-modal and dynamic roles

In OC, individuals 'take' or 'make' different roles which shape the products and goods that are created within OC as well as the adoption and usage of them. Due to loose coordination of OC, such roles are usually not assigned through formal governance but are achieved, emerge, or are even enacted in the moment in

UNCORRECTED PROOF

Please cite this article in press as: S. Brunswicker et al., Big Data for Open Digital Innovation – A Research Roadmap, Big Data Research (2015), http://dx.doi.org/10.1016/j.bdr.2015.01.008

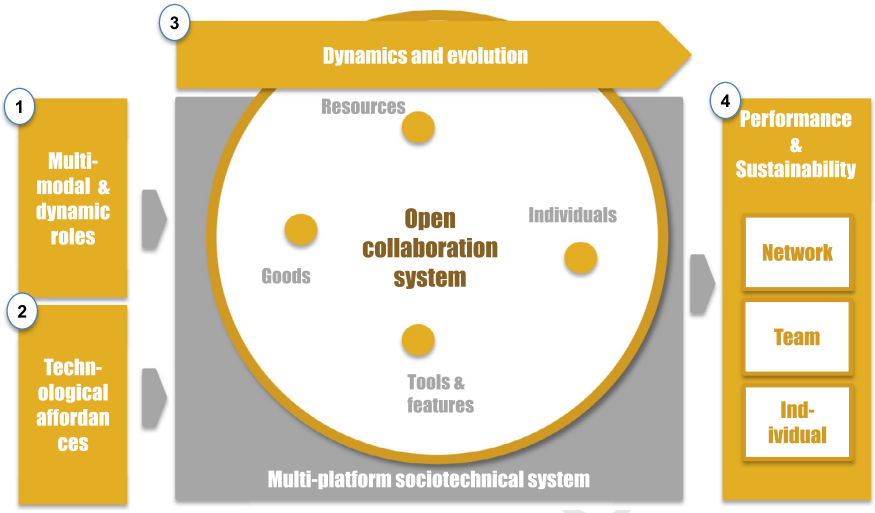

Fig. 1. Research framework.

a self-defined manner. In addition, individuals can hold multiple roles and may change roles over time [13,29]. Existing literature proposes a range of different roles, and also takes different theoretical lenses to identify and explain them. Examples of roles are functional leaders, who contribute to a significant degree to a project and lead by example and be determining the nature of the project [29], 'free riders', who hardly contribute to the development but only want to benefit from it [2], to boundary spanners who span across different subgroups [14], or service function roles like mediators or organizers [13]. Which roles matter for collaboration, knowledge creation, and decision making, and how they are achieved or enacted, is of utmost important to understand the inner working of OC. However, it is controversially discussed. There is no generalizable way of modeling them and the particular nature of the OC system – as a dynamic sociotechnical system – is not sufficiently considered. For example, network theory offers one theoretical lens to identify roles and collaborative behavior in OC in a data-driven manner by focusing on the structural dimension between individuals [36]. This theory has become a popular theoretical tool for studying social networks and contagious processes like diffusion of rumor, disease, and also innovation. Following he proposed assumptions of social network theory and literature on social capital [18], researchers furnish a range of centrality measures, such as 'closeness centrality' or 'betweenness centrality' (e.g. [15,25,4], just to name a few), to describe how location in a network topology determines critical roles of an individual in a social network [8].

However, OC is more than just a social network in which individuals are connected through social ties. It is a multi-modal network connecting multiple nodes of people, their resources, goods, and the tools available [29] rather than a mere social network connecting people to people through social relationships [21,31]. In addition, it represents a dynamic network in which there are tipping points, and stochastic events (see e.g. [35]). The debate on different roles in OC is further intensified due to the particular nature of this organizational form. Participants are jointly embedded in different social and technical contexts and connected through multiple activities and practices, and thus, the notion of 'centrality' is significantly different from the original idea of social embeddedness [18]. It relates to complex goal-oriented interactions and practices that take place in digital sociotechnical environments. 'Centrality' co-exists at different levels, and evolves dynamically over time. To resolve this debate and to model roles through a network theoretical lens, we need to dynamically consider a range of attributes of the nodes of these multi-modal networks – the users, the goods, the resources, and the tools – and the graphs connecting them. These graphs represent different kinds of goal-oriented and tangible activities and practices. We also need to consider the particular context (e.g. the particular technology platform the user is embedded in) and 'subpopulations' in which these interactions occur.

In addition, we need to enrich network models with insights from innovation and behavioral theory which highlight that the nature of the good – whether it is a rival good or not – and the heterogeneity of the resources of the individuals within the network – as well as the individual's tendency towards cooperation is a critical variable in OC. Further, a purely network theoretical perspective is not sufficient to identify roles in OC and to explain and predict their effect on cooperative behavior on OC. We need to complement a 'global' network perspective, with a local-individual behavioral perspective, as role making implies that individuals enact roles in the moment. Thus, it is the local individual activities that characterize enacted roles in OC. Such roles emerge in response to tension fluctuations in OC. For example, Kane et al. [22] identified roles-including flitterer, idea champion, and defender – that participants in a Wikipedia article use for collaborating on an article. The flitterer is a participant who comes to the community, places an idea, and then leaves [22]. In OC such self-enacted roles may be critical to drive contribution. However, it requires novel data-driven research and pattern-oriented mining to identify these roles and generalize whether they drive collaboration. It requires the analysis of local behavioral patterns and activities 'in the moment', in a particular context and in relation to subpopulations.

Guiding research question(s). *How can we conceptualize and measure roles from a sociotechnical, multi-modal network, and multilevel point of view? How we can better capture dynamic 'enacted' roles and move beyond the structural perspective? How do different roles effect collaboration and new knowledge creation?*

Analytical challenges. In the lights of the abundance of data about individual behavior, there are novel opportunities to model and predict roles in OC. However, there are a range of conceptual and analytical challenges. First, we require new linked data to address the multi-modal character of OC as well as means to capture and measure emerging multi-dimensional practice capital. There are meta-models available to model complex networks and their dynamics for organizations. However, they do not reflect the self-organizing nature of OC as organizational form. We require novel dynamic network 'meta' models, and also consider the

Please cite this article in press as: S. Brunswicker et al., Big Data for Open Digital Innovation – A Research Roadmap, Big Data Research (2015), http://dx.doi.org/10.1016/j.bdr.2015.01.008

local pattern-oriented conceptualization of roles and how roles related to other subpopulation in the moment. In addition, for large datasets we need more efficient algorithms to identify roles via centrality measures.

2.2. Thematic area 2: technological affordances for collaboration and innovation

Existing OC environments make use of various tools and features including APIs, chat rooms, wikis, recommender systems (e.g. Amazon review), or visualization techniques (e.g. tag clouds, network visualization). Today, researchers have little empirical insights about their affordances for collaboration, knowledge exchange, knowledge creation, and innovation. Innovation scholars regularly treat the technology as a "blackbox" and do not consider the digital environment and the tools and features available within it. However, following recent claims of transformative scholars we need to take turn and acknowledge the constitutive role of technology [32]. Technological features yield novel affordances or constraints which can be transformative and extremely critical for complex problem solving and activating individuals to change and solve problems in areas like healthcare, science, and alike. The concept of affordances refers to an action potential, that is, to what an individual or organization with a particular purpose can do with a particular technology or information systems of "tool" in a cyberinfrastructure [17]. Technologies may also yield a constraint as they hold back individuals from accomplishing a particular goal when using a technology or system. Thus, affordances and constraints are distinct from features. Affordances are relational concepts and are different from features.

In OC systems we expect flexible affordances that are used in creating innovations characterized by convergence and generativity [37]. For example, an architecture affording knowledge evolution allows anyone to see what knowledge has been collectively created. Such transparency may encourage individuals to contribute. Use of knowledge evolution visualization has yielded knowledge exchange. As with Wikipedia, such an affordance may allow individuals who have little time to engage in a process to eventually evolve to a more integrative and comprehensive solution [27]. Features and tools like visualization tools or recommender systems may yield affordances for evolutionary knowledge contribution, or even constraints [22]. Indeed, there might be novel features required to yield evolutionary knowledge creation and innovation that sustain over time. Technology affordances and constraints in OC are quite particular as they often span different context, systems, and platforms, such as open data environments or science networks like NanoHub.org. In addition, features are not static but change over time as individual's make changes not just to goods but also features. For example, in NanoHub.org users build their own simulation tools, which become a novel artifact for other users to create novel knowledge. Indeed, a deeper understanding and more generalizable evidence on technological affordances and constraints is required to build and advance a research community around OC.

Guiding research question(s). *What technological affordances and constraints within OC systems emerge from novel technological features (such as visualization tools)? How do they affect the dynamics of collaboration and the ability of individuals to create really novel solutions and tackle complex problems?*

Analytical challenges. Studying affordances has always been a challenging research task as it requires researchers to have access to relational data. Indeed, the explanatory (or even predictive) power of affordances can only be actualized if two conditions are met: (1) the technology and the technological feature are clearly defined and (2) the domain of activity is limited to a specific set of activity [12]. Meeting these conditions requires micro-level behavioral data about technology-use relationships. At the same time, these data must cover multiple contexts in order to support drawing generalizable conclusions. A pattern-based approach towards computationally identifying patterns about affordances across large user data sets and subpopulations may offer new means to arrive at explanatory or even predictive conclusions. In addition, collecting new data in an experimental and action-oriented approach will offer new means to receive fine-grained data about technology-use relationships.

2.3. Thematic area 3: explaining and predicting performance and sustainability of OC

OC has drawn scholarly interest because of its potential performance impact on innovation and its implications for society. Innovation and performance relate to different levels: the individual's performance, the performance at goods, the overall system level. The impact at the industry level is also puzzling. Some firms have been affected positively, others negatively. The free encyclopedia Wikipedia has managed to achieve the quality of Encyclopedia Britannica, which, after 244 years in circulation, has ceased to produce their books. Apparently, OC is transforming industries. Recent modeling drawing upon innovation theory and the concept of private-collective innovation argue that with increasingly decreasing communication and production costs, open and collaborative innovation models will increasingly become superior to producer innovation models [2]. However, what affects the performance of OCs – even why they are viable – remains critical question. An agent-based model has been recently developed theorizing that the dynamics in the system has a significant impact on the ability to solve complex problems in an open and collaborative way [1, 5]. Performance also relates to the question of sustainability and continuous value creation for all participants of OC. Taking an evolutionary perspective we argue that the OC system will go through phases of tensions. In OC such tensions relate to the competition versus collaboration at the individual level, creative abrasion versus 'being a stranger', time required for contributions versus time constraints, flexibility versus standardization. Generative responses rather than structure are required to ensure sustainability and respond to these tensions [6]. The emergence of different roles and technological affordances that support fluidity and dynamics may enable OC system to overcome these tensions and to sustain itself. Understanding what the dynamic drivers of continuous innovation and sustainability of OC are is of utmost importance.

Guiding research question(s). *How does the multilevel design of OC (related to roles and technological features and affordances) affect the dynamics of the system? What models enable us to explain and predict the self-organizing nature and the sustainability of OC? What are the particular drivers that affect sustainability of OC at different levels, the individual, product (good), the group and the overall network level?*

Analytical challenges. Studying the evolutionary process of new knowledge creation and the sustainability of OC is challenging, whether it is at the individual, the team, or the overall organizational level. A purely structural approach is not sufficient. Due to lack of empirical data some organizational scholars have mostly relied on agent-based modeling and randomly generated data to run experiments. However, big data and breadcrumbs about behavioral data at the individual as well as the network level data offer novel means to better link such computational models to empirical data. We also need to reconsider existing computational models

UNCORRECTED PROOF

Please cite this article in press as: S. Brunswicker et al., Big Data for Open Digital Innovation – A Research Roadmap, Big Data Research (2015), http://dx.doi.org/10.1016/j.bdr.2015.01.008

that have been developed for studying organizational models, as these computational models do not reflect the particular dynamic and sociotechnical nature of OC. In addition, longitudinal and fine-grained data might offer us the ability to statistically examine the sustainability of the overall OC system. However, we require linked data about individuals, goods, tools, and their particular context, detailed attributes, insights about dynamics over time, and highly performant computational infrastructures.

3. Analytic tools and computational techniques

To address such major research questions, one needs to rely on different research methodologies, tools and computational models. In the following, we elaborate on these:

- **Meta-network modeling:** A meta-network is a multi-mode, multi-link, multi-level network [23,9]. This is of relevance for all the themes in our research roadmap, but in particular for themes 1 and 3. Such meta-networks are required to predict node (actor) role attributes, actions, and behavior (including collaboration and knowledge exchange) on the basis of network linkages. Scholars in computational organizational science have already developed some initial meta-network models [9]. However, while these models are appropriate for most organizational forms, we need to meta-network models able to tackle the specifics of OC and the multi-modal networks that form around digital goods, resources, tools, and technological platforms. One approach is to take a multi-level practice and 'capital' view which form a meta-network around different 'practices'.
- **Network discovery and network analysis methodologies:** An important research task for OC research is network discovery and network analysis. One approach is to take a multi-level practice 'capital' view to discover the network, by which collaborative work in sequential or simultaneous tasks is seen as a form of connectedness and embeddedness in the sociotechnical system [18]. Previous work on network affiliation methodology used in prior research on network analysis can be leveraged on and extended to include co-participation in different 'practice' levels. According to this approach, individuals who work together are considered to be connected directly proportional to their level of commitment/contribution to a common sub-practice/task and inversely proportional to the distance in time between their contributions [29]. Once the network of practice is mapped, entity discovery methodologies can be applied that rely on community and cluster partitioning strategies to determine the functional role a user performs [36]. However, more sophisticated network analysis tools are required. In particular, for the identification of 'enacted' roles in the moment, one needs methods that go beyond existing multi-model network discovery techniques, and pattern-oriented approaches for analyzing user-behavior.
- **Dynamic network analysis and statistical prediction:** Studying the dynamics is of utmost important. Dynamic network analysis is an emerging field that brings together social network analysis (SNA), and link analysis in network science. There are different streams of dynamic network analysis, one of them is a statistical one [9]. Dynamic network analysis takes into account the temporal and dynamic analysis of a network. Among various types of models proposed for statistical modeling and prediction of network data, two stand out to be most promising. The first type consists of exponential random graph models [33], in which ties among nodes are assumed to be random variables and dependencies among these random variables are further imposed. The second type consists of latent social space models [19], which postulate the existence of a latent space and further assume that ties as random variables are determined by the positions of actors on the latent social space. Both models can be used to statistically infer fundamental rules and patterns of a social network based from observed data. An important research direction is to use these methodologies in the context of OC and to further advance them for the purpose of developing proper prediction models dynamic multi-modal networks. As these types of roles have not been investigated in large sociotechnical systems where collaboration is not tied to social ties but to practice and activity ties, further advances to such prediction models are needed. In addition, it is crucial to identify new statistical models to identify emerging 'enacted' roles that are less structurally oriented but result from interactions between users, or user-goods contributions.
- **Agent-based simulation models:** To predict complex behavior in OC systems agent-based modeling techniques are crucial. Agent-based models provide theoretical leverage where the global patterns of interest are more than the aggregation of individual attributes, but at the same time, the emergent pattern cannot be understood without a bottom up dynamical model of the micro-foundations at the relational level [26]. Recently an agent-based model to investigate the performance of OC from a behavioral point of view has been proposed [24]. So far, however, such model has not been validated with empirical data and also it needs to be extended in order to capture the sociotechnical and dynamic dimension of OC over time. Recently, one of the founding scholars of the field of computational organizational science has developed a toolkit CONSTRUCT to link meta-models, text mining, network analytics and agent-based modeling [9]. While this toolkit provides novel means to advance to explain and predict complex diffusion processes in organizational system, it does not reflect the particular nature of open digital innovation in virtual environments. Open collaboration often implies non-transitive relationships, and is not just shaped by social relationships.
- **Behavioral sequencing techniques and genetic computation:** To explain the sociomaterial character of behavior in OC, and to provide insights into complex interactions between humans and technologies, pattern-oriented approaches are needed for studying the entanglement of human activities and digital technologies in OC. A behavioral sequencing analysis allows one to capture the temporal interactions between users and technologies at the micro-level and understand how particular events trigger behavior. In addition, it also allows one to consider similarities across users and among subpopulations. Recently, genetic computation has entered the field of organizational studies and innovation [16]. However, it has not been applied to the phenomenon of OC. One possible approach is to identify routine activities via low entropy measures, as well as highly dynamic activities via high entry measures to identify highly dynamic roles and patterns that are enacted in the moment.
- **Collaborative and automated coding tools for unstructured text data:** To understand the micro-foundations of human behavior, we need to move beyond log-files and also engage in analysis of dialog among individuals engaging OC that takes place in chat rooms, on pin walls, etc. Today, analyzing the evolution of dialog is a time consuming process. Usually, there is no predefined set of codes available as researchers often need to engage in inductive and axial coding to make sense of data. Learning-oriented algorithms are needed that would allow one to build a rich lexica and meta-data related to particular dialog themes [34] e.g. the phenomenon of perspective taking and distancing in knowledge creating dialogs.

UNCORRECTED PROOF

Please cite this article in press as: S. Brunswicker et al., Big Data for Open Digital Innovation – A Research Roadmap, Big Data Research (2015), http://dx.doi.org/10.1016/j.bdr.2015.01.008

Beyond these computational and statistical tools, it is also important to integrate additional tools such as survey tools, and visualization tools. The latter are particular important for research and social experiments on technological affordances.

4. Concluding remarks

This paper has outlined research directions in the field of open collaboration. The increasing availability of datasets documenting on-line collaborations is making possible to provide answers to many research questions on a quantitative basis. However, just making available such datasets is not enough. We need new network models, data transformation techniques, data linkage techniques, and novel pattern analysis techniques to make possible to extract meaningful knowledge from such data. Capturing fine-grained and high quality interaction and collaboration data is also another important challenge that needs to be tackled. As part of our future work, we plan to design and prototype novel cyberinfrastructure platforms addressing such needs.

Acknowledgements

The work reported in this paper has been partially supported by the Purdue Research Center for Open Digital Innovation, the Purdue Cyber Center, and by the National Science Foundation under grant 1244708.

References

[1] E. Almirall, R. Casadesus-Masanell, Open versus closed innovation: a model of discovery and divergence, Acad. Manag. Rev. 35 (1) (2010) 27–47, http://dx.doi.org/10.5465/AMR.2010.45577790.
[2] Carliss Baldwin, Eric von Hippel, Modeling a paradigm shift: from producer innovation to user and open collaborative innovation, Organ. Sci. 22 (6) (2011) 1399–1417, http://dx.doi.org/10.1287/orsc.1100.0618.
[3] Robert P. Bostrom, Saurabh Gupta, Dominic Thomas, A meta-theory for understanding information systems within sociotechnical systems, J. Manag. Inf. Syst. 26 (1) (2009) 17–48, http://dx.doi.org/10.2753/MIS0742-1222260102.
[4] B.C. Ritt, System-level motivating factors for collaboration on Wikipedia: a longitudinal network analysis, Master's thesis, retrieved from ProQuest Dissertations and Theses Database, 2011, AAI 1501175.
[5] Sabine Brunswicker, Esteve Almirall, Melissa Lee, Towards a new era of developing public policies: when is openness superior to closed policy innovation? ESADE Working Paper, Barcelona, 2013.
[6] S. Brunswicker, A. Majchrzak, Continued value creation from open data: an ecosystem perspective, Working Paper no. 1, Research Center for Open Digital Innovation (RCODI), West Lafayette, IN, USA, 2014.
[7] S. Brunswicker, Wim Vanhaverbeke, Open Innovation in Small and Medium-sized Enterprises (SMEs): external knowledge sourcing strategies and internal organizational facilitators, J. Small Bus. Manag. (2014), forthcoming.
[8] R.S. Burt, The network structure of social capital, Res. Organ. Behav. 22 (2000) 345–423.
[9] Kathleen M. Carley, Jana Diesner, Jeffrey Reminga, Maksim Tsvetovat, Toward an interoperable dynamic network analysis toolkit, Special Issue Clusters 43 (4) (2007) 1324–1347, http://dx.doi.org/10.1016/j.dss.2006.04.003.
[10] Albert Cherns, The principles of sociotechnical design, Hum. Relat. 29 (8) (1976) 783–792, http://dx.doi.org/10.1177/001872677602900806.
[11] H. Chesbrough, S. Brunswicker, A fad or a phenomenon? The adoption of open innovation practices in large firms, Res. Technol. Manage. 57 (2) (2014), in press.
[12] Jennifer Earl, Katrina Kimport, Digitally Enabled Social Change. Activism in the Internet Age, MIT Press, Cambridge, MA, 2011.
[13] S. Faraj, S.L. Jarvenpaa, A. Majchrzak, Knowledge collaboration in online communities, Organ. Sci. 22 (5) (2011) 1224–1239, http://dx.doi.org/10.1287/orsc.1100.0614.
[14] Lee Fleming, David M. Waguespack, Brokerage, boundary spanning, and leadership in open innovation communities, Organ. Sci. 18 (2) (2007) 165–180, http://dx.doi.org/10.1287/orsc.1060.0242.
[15] Linton C. Freeman, A set of measures of centrality based on betweenness, Sociometry 40 (1) (1977) 35, http://dx.doi.org/10.2307/3033543.
[16] J. Gaskin, N. Berente, K. Lyytinen, Y. Yoo, Toward generalizable sociomaterial inquiry: a computational approach for 'Zooming In and Out' of sociomaterial routines, MIS Q. (2014), in press.
[17] J.J. Gibson, The theory of affordances, in: R. Shaw, J. Bransford (Eds.), The Ecological Approach to Visual Perception, Lawrence Erlbaum Associates, Hillsdale, NJ, 1977, pp. 127–143.
[18] Mark Granovetter, Economic action and social structure: the problem of embeddedness, Am. J. Sociol. 91 (3) (1985) 481–510.
[19] Mark S. Handcock, Adrian E. Raftery, Jeremy M. Tantrum, Model-based clustering for social networks, J. R. Stat. Soc., Ser. A, Stat. Soc. 170 (2) (2007) 301–354, http://dx.doi.org/10.1111/j.1467-985X.2007.00471.x.
[20] Joachim Henkel, Simone Schöberl, Oliver Alexy, The emergence of openness: how and why firms adopt selective revealing in open innovation, Res. Policy (2014), http://dx.doi.org/10.1016/j.respol.2013.08.014, forthcoming.
[21] G.C. Kane, M. Alavi, Casting the net and multimodel networks, Inf. Syst. Res. 19 (2) (2008) 253–272.
[22] G.C. Kane, A. Majchrzak, J. Johnson, G. Chen, A longitudinal model of perspective making and perspective taking within fluid online collectives, in: Proc. Internat. Conf. Inform. Systems, AIS Electronic Library, 2009, Paper 10.
[23] Michael J. Lanham, Geoffrey P. Morgan, Kathleen M. Carley, Social network modeling and agent-based simulation in support of crisis de-escalation, IEEE Trans. Syst. Man Cybern. 44 (1) (2014) 103–110, http://dx.doi.org/10.1109/TSMCC.2012.2230255.
[24] Sheen S. Levine, Michael J. Prietula, Open collaboration for innovation: principles and performance, Organ. Sci. (2013), http://dx.doi.org/10.1287/orsc.2013.0872, 131230050407004.
[25] E.Y. Li, et al., Co-authorship networks and research impact: a social capital perspective, Res. Policy 42 (9) (2013) 1515–1530.
[26] To be completed.
[27] A. Majchrzak, A. Malhotra, Towards an information systems perspective and research agenda on crowdsourcing for innovation, J. Strateg. Inf. Syst. 22 (4) (2013) 257–268, http://dx.doi.org/10.1016/j.jsis.2013.07.004.
[28] A. Majchrzak, M.L. Markus, Technological affordances and constraints in Management Information Systems (MIS), in: E. Kessler (Ed.), Encyclopedia of Management Theory, Sage Publications, 2014.
[29] S. Matei, E. Bertino, M. Zhu, L. Si, B. Britt, A research agenda for the study of entropic social structural evolution, functional roles, adhocratic leadership styles, and credibility in online organizations and knowledge markets, in: S. Matei, E. Bertino (Eds.), Roles, Trust, and Reputation in Social Media Knowledge Markets: Theory and Methods, Springer, New York, 2014, http://kredible.net/in/research-agenda.
[30] S. Matei, E. Bertino (Eds.), Roles, Trust, and Reputation in Social Media Knowledge Markets: Theory and Methods, Springer, New York, 2014.
[31] P.R. Monge, N.S. Contractor, Theories of Communication Networks, 2003.
[32] Wanda J. Orlikowski, Susan V. Scott, Sociomateriality: challenging the separation of technology, work and organization, Acad. Manag. Ann. 2 (1) (2008) 433–474, http://dx.doi.org/10.1080/19416520802211644.
[33] T. Snijders, Philippa E. Pattison, Garry L. Robins, Mark S. Handcock, New specifications for exponential random graph models, Sociol. Method. 36 (1) (2006) 99–153, http://dx.doi.org/10.1111/j.1467-9531.2006.00176.x.
[34] Haridimos Tsoukas, A dialogical approach to the creation of new knowledge in organizations, Organ. Sci. 20 (6) (2009) 941–957, http://dx.doi.org/10.1287/orsc.1090.0435.
[35] Zeynep Tufekci, Christopher Wilson, Social media and the decision to participate in political protest: observations from Tahrir square, J. Commun. 62 (2) (2012) 363–379, http://dx.doi.org/10.1111/j.1460-2466.2012.01629.x.
[36] H.T. Welser, E. Gleave, D. Fisher, M. Smith, Visualizing the signatures of social roles in online discussion groups, J. Soc. Struct. 8 (2) (2007) 1–32.
[37] Youngjin Yoo, Richard J. Boland, Kalle Lyytinen, Ann Majchrzak, Organizing for innovation in the digitized world, Organ. Sci. 23 (5) (2012) 1398–1408, http://dx.doi.org/10.1287/orsc.1120.0771.

UNCORRECTED PROOF