Some assembly required: leveraging Web science to understand and enable team assembly

Noshir Contractor

References	**This article cites 54 articles, 16 of which can be accessed free** http://rsta.royalsocietypublishing.org/content/371/1987/201203 85.full.html#ref-list-1 **Article cited in:** http://rsta.royalsocietypublishing.org/content/371/1987/20120385.full. html#related-urls
Subject collections	Articles on similar topics can be found in the following collections human-computer interaction (16 articles)
Email alerting service	Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click **here**

To subscribe to *Phil. Trans. R. Soc. A* go to:
http://rsta.royalsocietypublishing.org/subscriptions

Research

CrossMark
click for updates

Author for correspondence:

Noshir Contractor
e-mail: nosh@northwestern.edu

Some assembly required: leveraging Web science to understand and enable team assembly

Noshir Contractor

Department of Industrial Engineering & Management Science, McCormick School of Engineering, Department of Communication Studies, School of Communication, and Department of Management & Organizations, Kellogg School of Management, Northwestern University, Evanston, IL 60208, USA

Recent advances on the Web have generated unprecedented opportunities for individuals around the world to assemble into teams. And yet, because of the Web, the nature of teams and how they are assembled has changed radically. Today, many teams are ad hoc, agile, distributed, transient entities that are assembled from a larger primordial network of relationships within virtual communities. These assemblages possess the potential to unleash the high levels of creativity and innovation necessary for productively addressing many of the daunting challenges confronting contemporary society. This article argues that Web science is particularly well suited to help us realize this potential by making a substantial interdisciplinary intellectual investment in (i) advancing theories that explain our socio-technical motivations to form teams, (ii) the development of new analytic methods and models to untangle the unique influences of these motivations on team assembly, (iii) harvesting, curating and leveraging the digital trace data offered by the Web to test our models, and (iv) implementing recommender systems that use insights gleaned from our richer theoretical understanding of the motivations that lead to effective team assembly.

Royal Society **Publishing**
Informing the science of the future

1. Team assembly

(a) Background

The most important and complex decisions in governments, businesses and universities are made in teams. To solve today's most critical social and intellectual problems, then, we need teams with the best possible configuration of people. And yet, as recent events such as the earthquake in Haiti, the tsunami in Japan and the global financial crises indicate, assembling effective teams at short notice is a daunting task, for both intellectual and logistical reasons. A central challenge, catalysed by the Web, is that the size and complexity of teams and how they operate has changed radically [1]. Even before the recent IT revolution, Hollywood and the construction engineering industry provided early historical examples of the emergence of ad hoc groups that brought together people with different skills from their latent networks for a specific task over a finite time period. Today, unfettered to their local confines and aided by a plethora of spigots providing information about potential collaborators, individuals exercise much greater autonomy in assembling teams. Furthermore, teams (such as firefighters) exercise much greater autonomy in linking with other teams (such as the police) to assemble multi-team systems (MTSs) in order to accomplish higher order goals (such as disaster response). Clearly, an increasing preponderance of contemporary teams—in social, political and economic contexts—are ad hoc, agile, flexible, transient entities that emerge from a larger primordial latent network of relationships for a short duration to accomplish a wide variety of tasks and then dissolve to be reconstituted only with a different configuration at some later point in time.

While there is an incipient awareness of how team collaborations can spearhead socio-economic change, we still have sparse socio-technical knowledge of how potentially globally distributed teams and systems of teams are assembled, or how a given mode of assembly impacts effectiveness. Our lack of understanding also hinders efforts at enabling the assembly of effective teams. This study seeks to address these two limitations by arguing that Web science offers the perfect interdisciplinary forum to develop a theoretical, methodological and computational framework to both understand and enable the socio-technical dynamics, shaping the assembly of teams in distributed global contexts. While there is a vibrant body of related research focused on developing theories and tools to enable and understand how teams collaborate, this study is more specifically focused on understanding a precursor to the collaboration itself—the mechanisms by which collaborations are assembled.

(b) Technology and the assembly of teams

Teams have long been classified according to the extent to which team members act autonomously [2]. At the low end of the autonomy continuum are manager-led teams whose members are assigned to come together to execute a task. Self-managing teams are somewhat more autonomous, determining their internal work processes, whereas self-designing teams are even more autonomous, assembling themselves and setting their own direction and evaluative standards. Finally, the most autonomous teams, self-governing, not only have authority for their assembly, internal processes and overall direction, but also have significant authority to assemble the relations between teams as they come together to form larger organizational entities.

The Web has fundamentally reshaped the autonomy associated with all of these types of teams. Even formally constituted teams (i.e. manager-led, self-managing) are becoming more autonomous as members tap into their 'project network' to augment their 'project team' [3]. Highly autonomous teams now have the capacity to self-organize across all types of barriers including cultural and geographical [4,5]. Importantly, people self-organize within the bounds of traditional organizations, but they also do so outside the boundaries of formal organizations. In fact, these informal, self-organizing collectives tackle problems, such as the development of open source software [6–8].

A relatively recent concept that better captures these more complex collectives is the MTS. Based on the assumption that no team works in a vacuum, researchers expand the view of team to include the system level at which inter-team process is one of the core determinants of MTS and team effectiveness [9–11]. Formally defined, an MTS consists of 'two or more teams that interface directly and interdependently in response to environmental contingencies toward the accomplishment of collective goals' [12, p. 290]. A key distinction between the team and MTS arises because of the multi-level goal hierarchy [11]. In teams, members work towards their self-interest and towards the goals of the team. However, in an MTS, individuals, in addition to their own motives, have at least two focal collective motives: the team and the MTS. Given that individuals have finite effort and attentional resources, they must often decide where to allocate their resources: do they work for the good of the team, or do they work for the good of the larger system [13]? For instance, while both the firefighting and police teams share a higher level goal of responding to a disaster, the police team's goal of protecting any incriminating evidence is often at odds with the firefighting team's desire to rescue a life even if it is at the cost of potentially destroying police evidence.

A major reason why the Web has fundamentally reshaped assembly of teams and MTSs is the ever-increasing availability of useful information about potential and current collaboration partners in social and organizational contexts. Prior to the advent of computing technologies, people relied on a limited range of personal communications and (when available) paper records to assess the competence and desirability of potential collaborators. The ability of computational systems to capture and collate large volumes of information about behaviour and performance gives the contemporary workforce far superior data and metrics to inform their decisions regarding assembly of teams and MTSs. In addition to these passive recommendations, we are also witnessing the emergence of more active recommendation systems. These systems include reputational rating applications, such as those provided by eBay and eLance, and recommender systems, analogous to those provided in e-commerce sites such as Amazon, but tailored to recommend particular individuals (e.g. Angie's List).

One relatively uncharted arena in which recommender systems are likely to have significant impact is the assembly of virtual teams. Despite the widespread prevalence of virtual teams, recruitment of members into virtual teams is often conducted on an ad hoc basis rather than by systematically exploring the expertise and networks among potential team members. Therefore, a key issue is how to assemble teams that are optimally effective according to various criteria such as effectiveness, resilience and cohesiveness. In research on teams, this has been historically reduced to the problem of group composition, which has been studied for several decades (for a review, see [14]). Generally speaking, bringing the best and brightest individuals together into the same team does not make it the highest performing team and sometimes fails to achieve objectives. Research has empirically demonstrated that team composition influences team performance. For example, gender composition [15], personality [16,17], teamwork knowledge [18], value and belief [19] and cognitive ability [20] are a few among many attributes that have been linked to team process and performance.

Over 25 years ago, Moreland [21] called for research on group formation that went beyond group composition that was focused on the attributes of the individuals rather than the links among them. He identified several types of social integration and posited their effects on group formation. Environmental integration influenced the assembly of groups when the physical, social or cultural environments provided the resources necessary for group formation. Behavioural integration explained assembly of groups when people became dependent on one another for the satisfaction of their needs. Affective integration produced groups when people developed shared feelings. And, finally, cognitive integration explained the emergence of groups when people became aware of shared personal characteristics. Moreland [21] advocated social networks as an approach to investigate the influence of these various forms of social integration. As discussed in §2, the Web has dramatically increased the influence of the network on team assembly in the past 25 years since then. Happily, the Web has also provided access to the digital

trace data that could be required by network analysts to explore how these different forms of social integration influenced team assembly.

2. Networks and the assembly of teams and multi-team systems

A new development, in the past decade, has been the growing recognition that teams operate within complex networks of teams, individuals and external organizations, and that these network structures—along with size and composition of the team—influence their effectiveness [22–26]. Hinds *et al.* [27] found that, when selecting future team members, people are biased towards others of the same race, others who have a reputation for being competent and hard-working, and others with whom they have developed strong working relationships in the past. These results suggest that people strive for predictability when choosing team members. Based on an investigation of close to 500 National Science Foundation research projects, Cummings & Kiesler [28] found that collaborating on a team was influenced by three factors: familiarity, geographical proximity and disciplinary similarity. They also found that working previously on a team influenced subsequent collaboration by helping overcome geographical and disciplinary hurdles to future team assembly. Research by Guimera *et al.* [29] from both artistic and scientific fields shows that successful teams have a higher fraction of incumbents (people who are not novices to the effort) when controlling for the size of the team. Teams that are less diverse typically have lower levels of performance. Further, scientific teams produce high output only within a narrow band of behaviour. Teams should have incumbents on them about 30–65% of the time and have repeated ties about 45–90% of the time. Outside these ranges, there is a distinct drop off of impact. Within, the 'bliss point' accounts for 75 per cent of the highest impact work, whereas, outside, this range accounts for about 80 per cent of the lowest impact work.

In addition to team assembly, empirical research also points to the importance of networks on the assembly of MTSs. Three recent studies demonstrate that team interaction processes have markedly different effects on team and MTS performance. Marks *et al.* [11] found that, when teams are highly interdependent, within-team processes are unrelated to performance. Rather, between-team processes predict MTS performance. Similarly, DeChurch & Marks [10] found between-team process explains additional variance in the performance of the MTS after controlling for the performance of the component teams. Going a step further, Davison *et al.* [30] found that directives given to MTSs to increase their overall amount of coordination within the team was actually detrimental to overall MTS performance. Instead, boundary spanning of key people was critical to success. The importance of networks to MTS performance is further supported by work on group social capital. Oh *et al.* [31] found that team outcomes are maximized when there are saturated, dense connections among the members of particular teams, but teams are also connected by bridging relationships. Importantly, Oh *et al.* [31] found an inverted U-shaped relationship between the number of cross-team ties and team-level performance.

These studies, although few in number, offer compelling evidence that the network structure of the team prior to its assembly has an indelible imprint on its effectiveness. Put another way, while the research on improving collaboration within teams has shown us strategies to make teams more effective, the networks shaping the assembly of the team provide boundary conditions on their effectiveness.

Given the incipient yet profoundly influential insights about the impact of the Web and networks on the assembly of teams and MTSs (hereafter just referred to as teams), the remainder of this study begins with an exploration of the multiple theoretical motivations for team assembly. Next, the study describes recent innovations in network analytic techniques that provide an opportunity to test multi-theoretical models (MTMs) of team assembly. Having theories and methods are necessary but not sufficient conditions for understanding team assembly. Hence, §3 explores the pivotal role of the Web in ushering the era of Big Data. Section 7 outlines the development of recommendation systems that leverage the insights obtained from our theoretically guided empirical analysis of the factors influencing team assembly.

3. Multi-theoretical multi-level model for team assembly

Over the past decade, we have developed a multi-theoretical multi-level (MTML) model to explain individuals' motivations to create, maintain, dissolve and reconstitute links with others in a network [32–35]. More recently, we have extended the model individuals' motivations to create, maintain, dissolve or reconstitute a team linkage with another individual [36]. The model explains these motivations on the basis of attributes of the individuals (such as skill, level, role, resources) as well as the extant links (such as communication, financial transactions, exchange of materials or services) among individuals within the network. Next, we discuss the eight families of MTML mechanisms illustrating their explanation for assembling teams:

(1) Theories of self-interest focus on how people make choices that maximize their individual utility function. That is, they are motivated to assemble in teams that enable them to seek the personal goals they wish to achieve. For instance, individuals will join a team that has members with skills they do not possess but might like to learn. Two primary theories in this area are the theory of social capital [37] and transaction cost economics [38].

(2) Theories of mutual interest and collective action examine how assembling teams produces collective outcomes unattainable by individual action. People assemble teams because they believe it serves their mutual interests in accomplishing common or complementary goals [39]. For instance, individuals might assemble into a team to contribute modest amounts of expertise or resources which when pooled together can help them accomplish an activity that was not possible by an individual.

(3) Contagion theories address questions pertaining to the spread of ideas, messages, attitudes and beliefs in the assembly of teams [40]. For instance, individuals may join a team simply because other members in the network have joined the team. That is, they were 'infected' by their network to engage in a contagious activity.

(4) Cognitive theories explore the roles that meaning, knowledge and perceptions play in assembly of teams. Grounded in transactive memory, decisions to assemble into teams are influenced by who or what people know [41]. For instance, an individual might choose to join a team because they 'think' that other team members have resources they may need. Or the individual might have resources to offer that they think members in the team might need.

(5) Exchange and dependency theories explain the assembly of teams on the basis of the distribution of information and material resources among network members [42]. People seek to assemble with those whose resources they need and who in turn seek resources they possess. For instance, an individual who needs information about a disaster response area may join a team with those who have resources to offer victims so that collectively they can exchange the resources they need with the resources they can offer.

(6) Homophily and proximity theories account for assembly of teams on the basis of trait similarity and similarity of place [43]. This would explain the study mentioned earlier [28] where members were more likely to assemble in teams with others sharing similar attributes or in geographical proximity.

(7) Balance theories [44,45] posit the assembly of teams being governed by a desire of consistency in one's relations. That is, individuals are more likely to assemble in teams with friends of their friends (transitive triads) rather than with strangers.

(8) Finally, coevolutionary theory posits that teams are typically assembled in the belief that they will increase overall fitness, measured as performance, survivability, adaptability and robustness [46]. Coevolutionary theory articulates how teams linked by intra- and inter-team networks compete and cooperate with each other for scarce resources [47–49]. For instance, members of one team assembled on the basis of homophily might observe that they are less effective than a competing team which used, say, an assembly mechanism based on exchange.
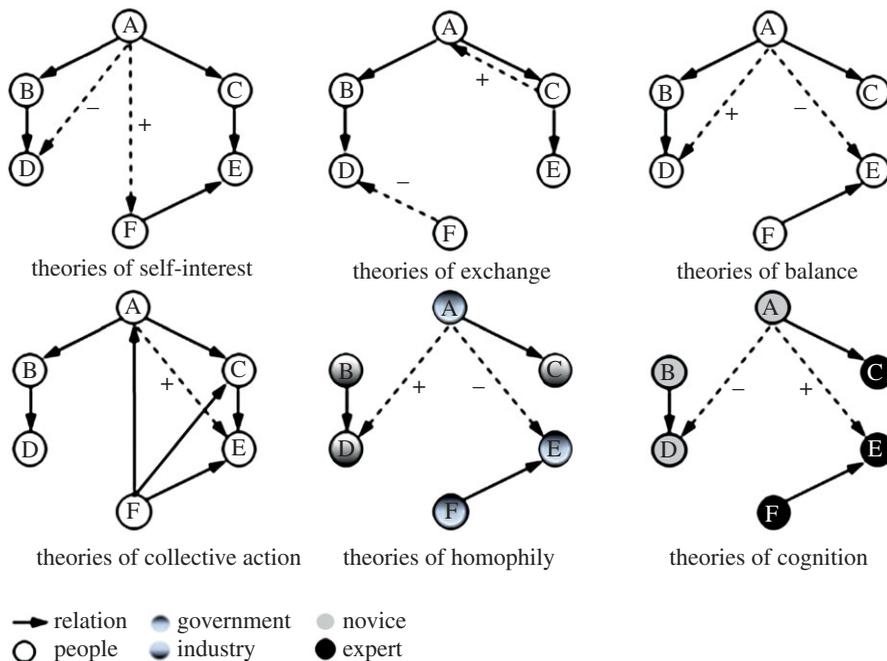
Figure 1. Network structural signatures association with team assembly. (Online version in colour.)

A key insight is that each of these assembly mechanisms has associated with it a distinct structural signature in the underlying network. Figure 1 shows the structural signatures associated with six of the eight assembly mechanisms described above. For instance, if a team were being assembled based on a logic of exchange, one would expect to see a larger number of reciprocated ties. Likewise, if the team were being assembled based on the logic of homophily, one would expect to see a larger number of ties between individuals sharing a common attribute (say, government employees) than those not sharing a common attribute (say, government versus industry employees). Of course, the premise behind an MTML model is that more than one mechanism might be contributing to the assembly of teams. In §4, we explore the challenges to statistically estimate network effects on team assembly and discuss a recent approach that has much promise.

4. Network analytic strategies for modelling the assembly of teams

The analyses of network data pose significant challenges that make the use of most traditional data analytic techniques inappropriate. As a result, most of the techniques used to compute inferential statistics in non-network analysis cannot be applied to network analysis. This is because a large proportion of inferential statistics used in non-network analysis make the assumption that the data are independently and identically distributed. Most standard statistical analyses that focus on attributes of (rather than relations among) individuals are premised on the assumption that the data are drawn from a distribution where the observations are independent. For instance, the height of an individual A does not have an impact on the height of an individual B. But network data observations are not independent of one another. That is, for instance, the presence of a communication tie between individual A and individual B could conceivably have an impact on the presence of a communication tie between individual A and some other individual C. Thus, many of the standard statistical techniques used to analyse attribute data are not appropriate for analysing network data.

Fortunately, recent advances in network analytic techniques make it possible to statistically unravel the extent to which any observed network has embodied in it one or more of these

structural signatures. These statistical techniques can be considered as the equivalent of a statistical MRI to discern the extent to which specific structural signatures are more (or less) likely to be observed in the network than one would expect by random chance. Exponential random graph models (ERGMs), also known as p^* models, have emerged as a prominent and promising effort to test complex network hypotheses [50,51]. In p^* models, the network data are represented as a large multi-graph with many types of actors and many relations. In addition, we have attribute information on the actors. We assume that these quantities are random variables 'tied together' by the theoretical concerns under study. As shown in figure 1, hypothesized team assembly mechanisms have unique structural signatures (for a review of all structural signatures, see [34]). For instance, if the assembly were motivated by balance, one would expect to observe more transitive triads than one would expect by chance. The general form of the class of (homogeneous) ERGMs is as follows:

$$P(\mathbf{X} = \mathbf{x}) = \kappa^{-1} \exp\left(\sum_{A \subseteq N_D} \lambda_A z_A(\mathbf{x})\right),$$

where (i) the summation is over structural signatures of types A; (ii) λ_A is the parameter corresponding to structural signatures of type A; (iii) $z_A(\mathbf{x})$ is the network statistic corresponding to structural signature A; and (iv) κ is a normalizing quantity to ensure that (1) is a proper probability distribution.

The model represents a probability distribution of graphs on a fixed node set, where the probability of observing a graph depends on the presence of the various structural signatures hypothesized in the model. One can interpret the structure of a typical graph in this distribution as the result of cumulating these particular structural signatures. The estimated parameters provide information about the presence of those structural signatures in the observed data [52]. In essence, ERGM/p^* analyses test the likelihood that the observed network was generated from the theoretically hypothesized structural signatures, thus reflecting the underlying social processes that lead to team assembly [51,53].

Clearly, techniques such as ERGM/p^* offer an unprecedented opportunity to test as well as generate new theoretical insights into the social motivations that explain how teams are assembled from the primordial networks in which they are embedded. However, until recently, one major hurdle remained: the lack of large-scale, high-resolution network data that would be required by these analytic tools in order to test the theoretical motivations outlined earlier. Section 5 describes how the Web, in addition to catalysing the creation of distributed teams, has also ushered in an era in which we now have the network data required to understand team assembly.

5. Harvesting the digital traces from the Web to model the assembly of teams

Until the last couple of decades, most network research relied on surveys and observations to collect network data. As a result, most studies analysed networks that were fairly small. Asking individuals to report on a survey their network ties with over 300 other people, for instance, would lead to what social scientists refer to as respondent fatigue. The same challenges would apply to individuals manually collecting observational network data. Clearly, traditional network data-collection approaches do not scale well. But the advent of the Web has substantially diminished these challenges and has ushered in a new era of 'computational social science' where for the first time social scientists can investigate and explore phenomena 'at scale' [25]. The Web is a treasure trove of archival data some of which are publicly available while others, for both privacy and commercial reasons, are proprietary. In the remainder of this section, we discuss our efforts to use data from the Web to understand the assembly of scientific teams working on research proposals and publications as well as the assembly of teams in massively multi-player online role-playing games.

(a) Assembly of scientific teams

Three studies [54–56] of about 20 million research articles recorded in the Web of Science database over five decades and an additional 2.1 million patent records over three decades found four important facts. First, for virtually all fields, research is increasingly done in teams. Second, teams typically produce more highly cited research than individuals do (accounting for self-citations), and this team advantage is increasing over time. Third, teams made up of people from different disciplines have an even higher impact than teams comprising individuals of a single discipline. And, fourth, teams made up of people from different disciplines at different geographical locations have even higher impact than teams comprising individuals at a single location. In what might seem a contradictory finding, in a study of projects funded by the US National Science Foundation, Cummings & Kiesler [9] report that projects which are more interdisciplinary and geographically distributed are less likely to be successful than those which are less interdisciplinary and geographically co-located. This seeming paradox is resolved when one recognizes that most interdisciplinary and geographically distributed teams are not very successful, but those that are succeed spectacularly. And it is only those success stories that appear as publications in the Web of Science. It is clear therefore that the assembly of a team has a strong bearing on its performance. But there is clearly more than team members' disciplinary background and geographical location that explain the variation in their success. Findings such as these have spurred a burgeoning interest in understanding 'team science' [57]. This field of inquiry, sometimes referred to as the Science of Team Science [58], is particularly focused on trying to explain why some teams are more effective than others based on their affiliations (reflected, for instance, by links among their websites), prior collaborations (reflected, for instance, by co-authorship gleaned from bibliometric databases), and common interests (reflected, for instance, by similar use of concepts obtained by keywords in publications or by text mining their documents). The emergence of this area of scholarship was directly influenced by the availability of digital data on the Web. Indeed, the momentum in this field is on the verge of a tipping point because of a major commitment by the US National Institutes of Health to fund a national research networking initiative that is making much of these data available using the linked open data standards [59].

A recent study conducted by our team [60] investigated the assembly mechanisms on a sample of 1103 successful and unsuccessful grant proposals submitted to two US National Science Foundation interdisciplinary programmes during a recent 3 year period. Our results indicate that individuals have a greater likelihood of assembling a proposal team with those with whom they have high levels of prior co-authorship and citation relationships. However, when we analysed separately the assembly mechanisms for successful and unsuccessful proposals, we found that individuals have a greater likelihood of assembling a successful proposal team with those with whom they have high levels of prior co-authorship, but low levels of prior citation relationships. This suggests that successful team outcomes are related to individuals who have developed a good working relationship (higher levels of prior co-authorship) but continue to draw upon different intellectual communities (hence the lower level of citation relationships).

(b) Assembly of teams in massively multi-player online role-playing games

A major barrier to understanding and explaining the role of the Web in team assembly is finding a suitable research environment. This environment would be one in which geographically distributed individuals are assembling in teams of varying sizes to accomplish a variety of tasks over varying durations. It would be one in which their actions, interactions and transactions are captured with precise time-stamps. Their individual and team outcomes would be recorded with well-defined metrics.

Massive multi-player online role-playing games (MMORPGs) offer a research environment that meets all of these requirements. MMORPGs have over 1.185 billion registered virtual world accounts and generated approximately $22.5 billion revenue in 2011. Players choose a character

(such as a scout, mage or a priest) and assemble into teams in order to accomplish quests and raids to kill monsters or obtain other in-game resources. While these games vary in theme and setting, they all entail individuals taking on complex tasks for which they often need to assemble into teams with others who have complementary skills. As a result, in an article in *Harvard Business Review*, Reeves *et al.* [61, p. 60] observed that, 'despite their fantasy settings, these online play worlds—sometimes given the infelicitous moniker MMORPGs (for "massively multiplayer online role-playing games")—in many ways resemble the coming environment. . . and thus open a window onto the future of real-world business leadership'. Hence, they argue that these online environments are in fact the ideal 'online laboratories' to understand and enable how we will use the Web to assemble into teams in the foreseeable future.

Over the past 5 years, we have built partnerships with the developers of several major MMORPGs in the USA, Europe and China. These partnerships have given us access to anonymized server logs maintained by the developers of these games. These logs chronicle every single action (such as making an in-game product such as a weapon or a potion), interaction (such as who is chatting with whom or broadcasting a message to a team) and transaction (such as buying a weapon or gifting a wardrobe item to another player) carried out in the game. These digital traces' size can be extremely large. Our team is working with a dataset that is over 40 terabytes! Here, again, we face challenges in developing algorithms to extract network relationships accurately from the corpus of digital traces [62].

This research has provided new insights into why people choose to team up with specific other individuals and with what implications for their performance. Our recent research [63], using the ERGM/p^* modelling techniques described earlier, indicates that team assembly is strongly influenced by theories of balance (we assemble into teams with friends of our friends), homophily (we assemble into teams with others of similar age and experience within the game), proximity (we are 22.6 times more likely to assemble into teams with someone who is within 50 km of us offline than those between 50 and 800 km). Curiously, we found a lack of evidence for gender homophily. Upon further analysis, we discovered that homophily was supported for men who liked to assemble into teams with other men. But homophily was not supported among women who also liked to assemble into teams with men! Ethnographic research helped us solve this puzzle. While many women play in these MMORPGs because they enjoy it, our ethnographic research indicated that several women play these games simply because it is their only chance to engage with their male significant others who are addicted to the game. As a result, rather than play with others of their own gender, women were more likely to assemble into teams with their male significant others.

Taken together, the theoretical models, analytic techniques and empirical data discussed thus far in this study have the potential to generate substantial new insights into the factors influencing the assembly of teams in general, and the assembly of effective teams in particular. The final step would be to leverage our understanding of team assembly to help enable more effective team assembly. Recommendation systems, discussed in §6, serve as a vehicle to implement theoretically guided and empirically validated team assembly strategies.

6. Enabling effective team assembly using recommender systems

Advances in the Web have made substantial progress in enabling synchronous as well as asynchronous collaboration tools. Unlike previous decades where the options for potential team members were limited, the ability to communicate and collaborate with anyone, anytime, have opened a plethora of new opportunities. While the Web makes it possible for us to communicate and collaborate with anyone, we are still exploring Web-based tools that will help us identify with *whom* we might want to communicate and collaborate. Efforts in this area have focused on the development of recommender systems [64]. While most systems focus on making recommendations about books, movies and other products/services, there has been a small but growing interest in the implementation of systems that make recommendations about connecting with other people. Some of these systems make recommendations based on attributes of the

people who are recommended. For instance, a person might be recommended because they have the requisite expertise [65,66]. For instance, Cosley *et al.* [67] have developed tools that examine a user's interests to recommend potential contributions they can make to community-based resources, such as Wikipedia. While this might work when recommending a user to a Wikipedia entry, it might not be effective when recommending a user to another person.

Algorithms based solely on expertise detection are not likely to be effective, because there is compelling evidence that people are not motivated to forge ties with others based solely on their expertise. An article in *Harvard Business Review* titled 'Competent jerks, lovable fools, and the formation of social networks' provides compelling evidence [68]. Some systems recommend people based on looking for people similar to the user. In the context of team assembly, similarity may not be the primary or even secondary goal; diversity in expertise and background and other social factors influence assembly of effective teams. Ounnas *et al.* [69] developed algorithms for recommending groups in undergraduate classes leveraging semantics to match people with similar interests while diversifying team composition based on the roles that individuals are inclined to play in groups (e.g. leader, implementer, etc.).

In response to these concerns, some recommender systems have incorporated a social component, which incorporates information from one's social network in making these recommendations [70–73]. Golbeck and others [74–76] have leveraged social relationships to make recommendations by inferring trust among users. In general, contemporary recommendation systems do not implement algorithms that take into account the semantics that characterize an individual's expertise, nor the social network assembly mechanisms that tie the user to a potential recommendation. This limitation is not surprising, because it has been very difficult to gain access to users' semantic and social network data.

However, the recent explosion in linked open data has unleashed new opportunities for developing a new generation of social network-enriched recommender systems [77]. These network data will enable recommender systems to use the parameters estimated in the statistical models for effective team assembly described in §4 to make recommendations for team assembly. From a technical standpoint, one can think of this effort as an exercise in link prediction. Link prediction models assume that the observed network has incomplete data (missing links) and therefore use techniques to predict the presence of links that are in the actual network but may not have been observed. Applying these techniques to an observed network that has complete data will result in the prediction of links that are not missing but ought to be there. These predicted links are, in essence, the recommended links. Recent reviews suggest that link prediction models fall into three categories: node-wise similarity, network topology such as the Katz method, and probabilistic models such as the relational Bayesian methods [78,79]. Preliminary results [80] indicate that recommendations made using ERGM/p^* models performed superior to models based on node-wise similarity, network topology (using the Katz method) and probabilistic models (using the relational Bayesian methods). This is particularly encouraging because it demonstrates that models grounded in social science theory about why we assemble in effective teams with others are superior to recommendations based on techniques that are driven purely by statistical models bereft of theoretical grounding.

7. Conclusion

This article has attempted to argue that Web science is particularly well suited to help us both understand and enable the assembly of teams—a challenge of paramount importance in helping us address societal challenges. This is a particularly opportune moment for Web science to serve a catalytic role because it offers the ideal incubator for significant interdisciplinary intellectual investment in four areas. First, as argued earlier, it offers an unprecedented opportunity to leverage—and extend—the corpus of multiple social scientific theories, operating at multiple levels, that contribute to the assembly of effective teams. Second, the specification of these MTML models has motivated the development of new network analytic methods and models (such as ERGM/p^*) to untangle the structural signatures reflecting the contributions of each of the

multiple theories in explaining the assembly of effective teams. Third, the presence of theories and methods would be ineffectual if we did not have access to a large volume of data to test these theories. Here, again, the Web offers an invaluable opportunity to advance our skills in harvesting, curating and manipulating the digital trace data in order to empirically test our theoretical models of team assembly. The potential of using digital trace data was illustrated by investigating the assembly of scientific teams and those playing massively multi-player online games. Finally, we discussed how recommender systems enable us to leverage our understanding of effective team assembly and turn those insights (or, more specifically, estimates from statistical models of effective team assembly) into helping enable more effective teams. Our preliminary findings indicate that link predictions (for recommendations) based on models grounded in social science theory outperform traditional link prediction models based purely on computational approaches. This augurs well for a more enlightened form of interdisciplinary efforts, embodied in the principles of Web science, as we address the grand societal challenges of the twenty-first century.

References

1. Zammuto RF, Griffith TL, Majchrzak A, Dougherty DJ, Faraj S. 2007 Information technology and the changing fabric of organization. *Organ. Sci.* **18**, 749–762. (doi:10.1287/orsc.1070.0307)
2. Hackman JR. 1987 The design of work teams. In *Handbook of organizational behavior* (ed. JW Lorsch), pp. 315–342. Englewood Cliffs, NJ: Prentice Hall.
3. Cummings JN, Pletcher C. 2011 Why project networks beat project teams? *MIT Sloan Manage. Rev.* **52**, 75–80.
4. Beyene T, Hinds PJ, Cramton C. 2011 *Walking through jelly: language proficiency, emotions, and disrupted collaboration in global work*. Alston, MA: Harvard Business School.
5. Hinds P, Liu L, Lyon J. 2011 Putting the global in global work: an intercultural lens on the practice of cross-national collaboration. *Acad. Manage. Ann.* **5**, 135–188. (doi:10.1080/19416520.2011.586108)
6. Hahn J, Moon JY, Zhang C. 2008 Emergence of new project teams from open source software developer networks: impact of prior collaboration ties. *Inform. Syst. Res.* **19**, 369–391. (doi:10.1287/isre.1080.0192)
7. Crowston K, Wei K, Howison J, Wiggins A. 2012 Free/libre open source software: what we know and what we do not know. *ACM Comput. Surv.* **44**, 7. (doi:10.1145/2089125.2089127)
8. Shen C, Monge P. 2011 Who connects with whom? A social network analysis of an online open source software community. *First Monday* **16**, 6–6.
9. Cummings J, Kiesler S. 2005 Collaborative research across disciplinary and organizational boundaries. *Soc. Stud. Sci.* **35**, 703–722. (doi:10.1177/0306312705055535)
10. DeChurch LA, Marks MA. 2006 Leadership in multiteam systems. *J. Appl. Psychol.* **91**, 311–329. (doi:10.1037/0021-9010.91.2.311)
11. Marks MA, DeChurch LA, Mathieu JE, Panzer FJ, Alonso A. 2005 Teamwork in multi-team systems. *J. Appl. Psychol.* **9**, 964–971. (doi:10.1037/0021-9010.90.5.964)
12. Mathieu JE, Marks MA, Zaccaro SJ. 2001 Multi-team systems. In *Handbook of industrial, work, and organizational psychology*, vol. 2 (eds N Anderson, D Ones, HK Sinangil, C Viswesvaran), pp. 289–313. London, UK: Sage Publications.
13. Kanfer R, Kerry M. 2011 Motivation in multiteam systems. In *Multiteam systems: an organizational form for dynamic and complex environments* (eds SJ Zaccaro, MA Marks, LA DeChurch), pp. 81–108. New York, NY: Taylor & Francis.
14. Moreland RL, Levine JM, Wingert ML. 1996 Creating the ideal group: composition effects at work. *Underst. Group Behav.* **2**, 11–35.
15. LePine JA, Hollenbeck JR, Ilgen DR, Colquitt JA, Ellis A. 2002 Gender composition, situational strength, and team decision-making accuracy: a criterion decomposition approach. *Organ. Behav. Hum. Decis. Process.* **88**, 445–475. (doi:10.1006/obhd.2001.2986)

16. Bell ST. 2007 Deep-level composition variables as predictors of team performance: a meta-analysis. *J. Appl. Psychol.* **92**, 595–615. (doi:10.1037/0021-9010.92.3.595)

17. Barrick MR, Stewart GL, Neubert MJ, Mount MK. 1998 Relating member ability and personality to work-team processes and team effectiveness. *J. Appl. Psychol.* **83**, 377–391. (doi:10.1037/0021-9010.83.3.377)

18. Morgeson FP, Reider MH, Campion MA. 2005 Selecting individuals in team settings: the importance of social skills, personality characteristics, and teamwork knowledge. *Pers. Psychol.* **58**, 583–611. (doi:10.1111/j.1744-6570.2005.655.x)

19. Harrison DA, Price KH, Gavin JH, Florey AT. 2002 Changing effects of surface- and deep-level diversity on group functioning. *Acad. Manage. J.* **45**, 1029–1045. (doi:10.2307/3069328)

20. LePine JA, Hollenbeck JR, Ilgen DR, Hedlund J. 1997 Effects of individual differences on the performance of hierarchical decision-making teams: much more than G. *J. Appl. Psychol.* **82**, 803–811. (doi:10.1037/0021-9010.82.5.803)

21. Moreland RL. 1987 Group formation. In *Group processes. Review of personality and social psychology*, vol. 8 (ed. C Hendrick), pp. 80–110. Thousand Oaks, CA: Sage Publications, Inc.

22. Contractor N, Su C. 2011 Understanding groups from a network perspective. In *Research methods for studying groups and teams: a guide to approaches, tools, and technologies* (eds A Hollingshead, MS Poole), pp. 284–310. New York, NY: Taylor & Francis/Routledge.

23. Faraj S, Johnson SL. 2011 Network exchange patterns in online communities. *Organ. Sci.* **22**, 1464–1480. (doi:10.1287/orsc.1100.0600)

24. Katz N, Lazer D, Arrow H, Contractor N. 2004 Network theory and small groups. *Small Group Res.* **35**, 307–332. (doi:10.1177/1046496404264941)

25. Lazer D *et al.* 2009 Computational social science. *Science* **323**, 721–723. (doi:10.1126/science.1167742)

26. Poole MS, Contractor NS. 2011 Conceptualizing the multiteam system as a system of networked groups. In *Multiteam systems: an organizational form for dynamic and complex environments* (eds SJ Zaccaro, MA Marks, LA DeChurch), pp. 193–224. New York, NY: Routledge Academic.

27. Hinds P, Carley KM, Krackhardt D, Wholey D. 2000 Choosing work group members: balancing similarity, competence, and familiarity. *Organ. Behav. Hum. Decis. Process.* **81**, 226–251. (doi:10.1006/obhd.1999.2875)

28. Cummings JN, Kiesler S. 2007 Coordination costs and project outcomes in multi-university collaborations. *Res. Policy* **36**, 1620–1634. (doi:10.1016/j.respol.2007.09.001)

29. Guimera R, Uzzi B, Spiro J, Amaral LAN. 2005 Team assembly mechanisms determine collaboration network structure and team performance. *Science* **308**, 697–702. (doi:10.1126/science.1106340)

30. Davison RB, Hollenbeck JR, Barnes CM, Sleesman DJ, Ilgen DR. 2012 Coordinated action in multiteam systems. *J. Appl. Psychol.* **97**, 808–824. (doi:10.1037/a0026682)

31. Oh H, Chung MH, Labianca G. 2004 Group social capital and group effectiveness: the role of informal socializing ties. *Acad. Manage. J.* **47**, 860–875. (doi:10.2307/20159627)

32. Contractor NS, Monge PR. 2002 Managing knowledge networks. *Manage. Commun. Q.* **16**, 249–258. (doi:10.1177/089331802237238)

33. Contractor NS, Monge PR. 2003 Using multi-theoretical multilevel models to study adversarial networks. In *Dynamic social network modeling and analysis: workshop summary and papers* (eds R Breiger, K Carley, P Pattison), pp. 324–344. Washington, DC: National Research Council.

34. Contractor N, Wasserman S, Faust K. 2006 Testing multi-theoretical multilevel hypotheses about organizational networks: an analytic framework and empirical example. *Acad. Manage. Rev.* **31**, 681–703. (doi:10.5465/AMR.2006.21318925)

35. Monge PR, Contractor NS. 2003 *Theories of communication networks*. Oxford, UK: Oxford University Press.

36. Contractor N, Su C. 2011 Understanding groups from a network perspective. In *Research methods for studying groups and teams: a guide to approaches, tools, and technologies* (eds A Hollingshead, MS Poole), pp. 284–310. New York, NY: Routledge.

37. Burt RS. 1992 *Structural holes: the social structure of competition*. Boston, MA: Harvard University Press.

38. Williamson OE. 1991 Comparative economic organization: the analysis of discrete structural alternatives. *Adm. Sci. Q.* **36**, 269–296. (doi:10.2307/2393356)

39. Fulk J, Heino R, Flanagin A, Monge P, Bar F. 2004 A test of the individual action model of organizational information commons. *Organ. Sci.* **15**, 569–585. (doi:10.1287/orsc.1040.0081)

40. Burt RS. 1987 Social contagion and innovation: cohesion versus structural equivalence. *Am. J. Soc.* **92**, 1287–1335. (doi:10.1086/228667)

41. Hollingshead AB, Fulk J, Monge P. 2002 Fostering intranet knowledge-sharing: an integration of transactive memory and public goods approaches. In *Distributed work: new research on working across distance using technology* (eds S Kiesler, P Hinds), pp. 235–255. Cambridge, MA: MIT Press.

42. Cook KS. 1982 Network structures from an exchange perspective. In *Social structure and network analysis* (eds PV Marsden, N Lin), pp. 177–218. Beverly Hills, CA: Sage.

43. McPherson JM, Smith-Lovin L. 1987 Homophily in voluntary organizations: status distance and the composition of face to face groups. *Am. Sociol. Rev.* **52**, 370–379. (doi:10.2307/2095356)

44. Heider F. 1958 *The psychology of interpersonal relations*. New York, NY: John Wiley.

45. Holland PW, Leinhardt S. 1975 The statistical analysis of local structure in social networks. In *Sociological methodology, 1976* (ed. DR Heise), pp. 1–45. San Francisco, CA: Jossey Bass.

46. Campbell DT. 1986 An organizational interpretation of evolution. In *Evolution at a crossroads: the new biology and the new philosophy of science* (eds DD Depew, BH Weber), pp. 33–167. Cambridge, MA: MIT Press.

47. Baum JAC. 1999 Whole-part coevolutionary competition in organizations. In *Variations in organization science: in honor of Donald T. Campbell* (eds JAC Baum, B McKelvey), pp. 113–135. Thousand Oaks, CA: Sage.

48. Bryant JA, Monge P. 2008 The evolution of the children's television community, 1953–2003. *Int. J. Commun.* **2**, 160–192.

49. Monge P, Heiss BM, Margolin DB. 2008 Communication network evolution in organizational communities. *Commun. Theory* **18**, 449–477. (doi:10.1111/j.1468-2885.2008.00330.x)

50. Frank O, Strauss D. 1986 Markov graphs. *J. Am. Stat. Assoc.* **81**, 832–842. (doi:10.1080/01621459.1986.10478342)

51. Robins G, Pattison P, Kalish Y, Lusher D. 2007 An introduction to exponential random graph (p*) models for social networks. *Soc. Netw.* **29**, 173–191. (doi:10.1016/j.socnet.2006.08.002)

52. Robins G, Snijders T, Wang P, Handcock M, Pattison P. 2007 Recent developments in exponential random graph (p*) models for social networks. *Soc. Netw.* **29**, 192–215. (doi:10.1016/j.socnet.2006.08.003)

53. Shumate M, Palazzolo ET. 2010 Exponential random graph (p*) models as a method for social network analysis in communication research. *Commun. Methods Measures* **4**, 341–371. (doi:10.1080/19312458.2010.527869)

54. Jones BF, Wuchty S, Uzzi B. 2008 Multi-university research teams: shifting impact, georgraphy, and stratification in science. *Science* **322**, 1259–1262. (doi:10.1126/science.1158357)

55. Wuchty S, Jones B, Uzzi B. 2007 The increasing dominance of teams in production of knowledge. *Science* **316**, 1036–1039. (doi:10.1126/science.1136099)

56. Wuchty S, Jones B, Uzzi B. 2007 Science commentary: why do team-authored papers get cited more? *Science* **317**, 1496–1498. (doi:10.1126/science.317.5844.1496b)

57. Börner K *et al.* 2010 A multi-level systems perspective for the science of team science. *Sci. Transl. Med.* **2**, 49cm24. (doi:10.1126/scitranslmed.3001399)

58. Falk-Krzesinski HJ *et al.* 2010 Advancing the science of team science. *Clin. Transl. Sci.* **3**, 263–266. (doi:10.1111/j.1752-8062.2010.00223.x)

59. Weber GM *et al.* 2011 Direct2Experts: a pilot national network to demonstrate interoperability among research-networking platforms. *J. Am. Med. Inf. Assoc.* **18**(Suppl. 1), i157–i160. (doi:10.1136/amiajnl-2011-000200)

60. Contractor N, Lungeanu A, Huang Y. 2011 Understanding the assembly of interdisciplinary scientific teams. In *Proc. of Conf. Innovation, Organizations, and Society: A Lambert Family Communication Conference, University of Chicago. Chicago, IL, 14–15 October 2011.*

61. Reeves B, Malone TW, Driscoll TO. 2008 Leadership's online labs. *Harv. Bus. Rev.* **86**, 58–66.

62. Williams D, Contractor N, Poole MS, Srivastava J, Cai D. 2011 The virtual worlds exploratorium: using large-scale data and computational techniques for communication research. *Commun. Methods Measures* **5**, 163–180. (doi:10.1080/19312458.2011.568373)

63. Huang Y *et al.* In press. Distance matters: exploring proximity and homophily in virtual world networks. *Decis. Support Syst.* (doi:10.1016/j.dss.2013.01.006)

64. Adomavicius G, Tuzhilin A. 2005 Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Trans. Knowledge Data Eng.* **17**, 734–749. (doi:10.1109/TKDE.2005.99)

65. Balog K, Azzopardi L, de Rijke M. 2006 Formal models for expert finding in enterprise corpora. In *Proc. 29th Annu. Int. ACM SIGIR Conf. on Research and Development in Information Retrieval, Seattle, Washington, 6–11 August 2006*. New York, NY: Association for Computing Machinery.

66. Zhang J, Khopkar T, Hansen DL. 2013 Recommender systems and expert locators. In *Encyclopedia of library and information sciences*, 3rd edn, pp. 4433–4441. Boca Raton, FL: Taylor & Francis.

67. Cosley D, Frankowski L, Terveen L, Riedl J. 2007 SuggestBot: using intelligent task routing to help people find work in wikipedia. In *Proc. 2007 Int. Conf. on Intelligent User Interfaces, Honolulu, HI, 28–31 January 2007*, pp. 32–41. New York, NY: Association for Computing Machinery.

68. Casciaro T, Lobo MS. 2005 Competent jerks, lovable fools, and the formation of social networks. *Harv. Bus. Rev.* **83**, 92–99, 149.

69. Ounnas A, Davis H, Millard D. 2008 A framework for semantic group formation. In *Proc. 8th IEEE Int. Conf. on Advanced Learning Technologies, Santander, Cantabria, Spain, 1–5 July*. New York, NY: IEEE.

70. Huang Y, Contractor N, Yao Y. 2008 CI-KNOW: recommendation based on social networks. In *Proc. 2008 Int. Conf. on Digital Government Research, Montreal, QC, 18–21 May 2008*. Digital Government Society of North America.

71. Kautz H, Selman B, Shah M. 1997 Referral web: combining social networks and collaborative filtering. *Commun. ACM* **40**, 63–65. (doi:10.1145/245108.245123)

72. McDonald DW. 2003 Recommending collaboration with social networks: a comparative evaluation. In *Proc. the SIGCHI Conf. Human Factors in Computing Systems, Ft. Lauderdale, FL, 5–10 April 2003*. New York, NY: Association for Computing Machinery.

73. Schafer JB, Frankowski D, Herlocker J, Sen S. 2007 Collaborative filtering recommender systems. In *The adaptive web* (eds B Peter, K Alfred, N Wolfgang). Berlin, Germany: Springer.

74. Golbeck J. 2008 Weaving a web of trust. *Science* **321**, 1640–1641. (doi:10.1126/science.1163357)

75. Ziegler C-N, Lausen G. 2005 Propagation models for trust and distrust in social networks. *Inf. Syst. Front.* **7**, 337–358. (doi:10.1007/s10796-005-4807-3)

76. Massa P, Avesani P. 2007 Trust-aware recommender systems. In *Proc. of the 2007 ACM Conf. on Recommender Systems (RecSys 2007), Minneapolis, MN, 19–20 October 2007*. New York, NY: Association for Computing Machinery.

77. Stankovic M, Wagner C, Jovanovic J, Laublet P. 2010 Looking for experts? What can linked data do for you? In *Proc. of the WWW2010 Workshop on Linked Data on the Web (LDOW 2010), Raleigh, NC, 27 April 2010*. CEUR Workshop Proceedings 628.

78. Liben-Nowell D, Kleinberg J. 2007 The link-prediction problem for social networks. *J. Am. Soc. Inf. Sci. Technol.* **58**, 1019–1031. (doi:10.1002/asi.20591)

79. Lü L, Zhou T. 2011 Link prediction in complex networks: a survey. *Physica A* **390**, 1150–1170. (doi:10.1016/j.physa.2010.11.027)

80. Huang Y, Zhang C, Fazel-Zarandi M, Devlin H, Lungeanu A, Wasserman S, Contractor N. 2012 Comparing efficacy of link prediction models for expert recommender systems. In *Proc. of the Int. Sunbelt Social Network Conference (Sunbelt XXXII), Redondo Beach, CA, 12–18 March 2012*.