COMMUNICATION METHODS AND MEASURES
https://doi.org/10.1080/19312458.2018.1430756

# How Team Interlock Ecosystems Shape the Assembly of Scientific Teams: A Hypergraph Approach

Alina Lungeanu [a], Dorothy R. Carter[b], Leslie A. DeChurch[c], and Noshir S. Contractor[c]

[a]Department of Communication Studies, Northwestern University, Evanston, USA; [b]Department of Psychology, University of Georgia, Athens, USA; [c]Department of Communication Studies, Northwestern University, Evanston, USA

**ABSTRACT**

Today's most pressing scientific problems necessitate scientific teamwork; the increasing complexity and specialization of knowledge render "lone geniuses" ill-equipped to make high-impact scientific breakthroughs. Social network research has begun to explore the factors that promote the *assembly* of scientific teams. However, this work has been limited by network approaches centered conceptually and analytically on "nodes as people," or "nodes as teams." In this article, we develop a "*team-interlock ecosystem*" conceptualization of collaborative environments within which new scientific teams, or other creative team-based enterprises, assemble. Team interlock ecosystems comprise teams linked to one another through overlapping memberships and/or overlapping knowledge domains. They depict teams, people, *and* knowledge sets as nodes, and thus, present both conceptual advantages as well as methodological challenges. Conceptually, team interlock ecosystems invite novel questions about how the structural characteristics of embedding ecosystems serve as the primordial soup from which new teams assemble. Methodologically, however, studying ecosystems requires the use of more advanced analytics that correspond to the inherently multilevel phenomenon of scientists nested within multiple teams. To address these methodological challenges, we advance the use of *hypergraph methodologies* combined with bibliometric data and simulation-based approaches to test hypotheses related to the ecosystem drivers of team assembly.

## Introduction

The idea that high-impact scientific breakthroughs are the work of "lone geniuses" has long lost its appeal. Narrative (e.g., Charney, 2003) as well quantitative accounts (Uzzi & Spiro, 2005) reveal that today's most pressing scientific problems—those within domains ranging from translational medicine to environmental sustainability, from cyber learning to disaster response—present a degree of complexity that necessitates scientific *teamwork* (Barabási, 2005; Wuchty, Jones, & Uzzi, 2007).

For the most part, studies of scientific teamwork have focused on the factors related to team effectiveness occurring *after* a team has assembled (e.g., National Research Council, 2015). However, studies of team functioning post-assembly miss important dynamics occurring *prior* to team assembly that impinge on scientists' decisions to join teams in the first place. Scientific organizations, such as universities or research institutes, afford individuals substantial autonomy and flexibility on forming or join new teams. Given that science is a human endeavor, scientists are susceptible to

**CONTACT** Alina Lungeanu alina.lungeanu1@northwestern.edu Northwestern University, Department of Communication Studies, Evanston, Illinois.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/hcms.

Supplemental data for this article can be access on the publisher's website.

natural human social preferences (e.g., McPherson, Smith-Lovin, & Cook, 2001) and cognitive limitations (e.g., De Solla Price, 1965; Dunbar, 1992) when making decisions related to team assembly. These natural human proclivities often lead individuals to assemble into teams that are suboptimal in their effectiveness. In fact, effectively assembling scientific teams can be a daunting task, both logistically as well as technically, and many scientific teams suffer from the consequences of suboptimal team assembly (Cummings & Kiesler, 2008).

To address these challenges, researchers have begun to investigate the mechanisms that promote the assembly of new scientific teams. Research on scientific team assembly often relies on *bibliometric* approaches, which leverage publically available information on publication, co-authorship, and/or citation activity. These approaches are used to understand how prior patterns of collaborative activities influence future team assembly. For the most part, research investigating drivers of scientific collaboration has depicted collaborations as ties between *pairs* of researchers (e.g., Cummings & Kiesler, 2007; Lungeanu, Huang, & Contractor, 2014). These "person-to-person" approaches focus on the likelihood of a collaboration tie between a pair of researchers in a team. Lost in these approaches, however, is the ability to distinguish a network of (1) three researchers linked pairwise because all three collaborated on one publication from (2) three researchers linked pairwise because pairs of them collaborated on three separate publications. Thus, the representation used by current network approaches that model collaboration as dyads are unable to discriminate between these two substantially different collaboration scenarios.

Scientists often work in teams including more than two individuals. Thus, representations of collaboration should explicitly characterize multiple individuals in teams. Furthermore, most scientists work on *multiple* teams, concurrently, and across time, as they engage with new and old collaborators to address research problems requiring unique and overlapping knowledge domains. Thus, representations of collaboration should explicitly characterize teams that overlap or "interlock" with other teams based on common members (i.e., *member interlocks*) and/or common research topics (i.e., *knowledge interlocks*). Poole and Contractor (2011) argue that, for the aforementioned reasons, we should examine how scientific teams nucleate within complex *multilevel ecosystems*. The structures of interlocking *teams* within scientific ecosystems are relevant to understanding team assembly because social phenomena, such as social bonding, knowledge generation, and learning, are team experiences that are likely to shape the assembly of future teams.

Research on scientific team assembly needs to move beyond a dyadic person-to-person framework that characterizes collaboration as a collection of pairs of researchers and explore the ways in which characteristics of the interlocking team structures, in which scientists are embedded, influence team assembly. This article leverages an extension to graph theory, *hypergraphs*, to address methodological limitations of current network approaches and better account for the nesting of individuals in teams and the patterns of interlocks among teams. We begin by describing key aspects of scientific team ecosystems, and operationalize these concepts using a hypergraph framework. Then, we provide exemplar hypotheses suggesting that there are certain characteristics of scientific ecosystems that enhance the likelihood that certain scientific teams will assemble. Finally, we demonstrate the combined use of hypergraph methodologies with an advanced computational technique to test these hypotheses. The hypergraph methodological approach enables researchers to operationalize the patterns of team interlock structures characterizing scientific ecosystems. The computational technique, *ecosystem simulations*, tests ecosystem-based hypotheses by comparing observed characteristics of scientific ecosystems with the ecosystem characteristics from a distribution of randomly generated ecosystems.

## Team interlock ecosystems

Research on scientific team assembly has significantly advanced our understanding of scientific teamwork by demonstrating that there are certain fundamental human tendencies that give rise to new teams. This research uncovered three categories of fundamental characteristics that predict future collaboration between any given *pair* of scientists: individual attributes of the scientists, prior collaboration relations between them, and characteristics associated with the broader structure in which the scientists' dyad is embedded.

Collaboration dyads are more likely to form when scientists have complementary skills (Lee & Bozeman, 2005), are geographically proximate (Cummings & Kiesler, 2007), and when both members have longer tenure, are affiliated with lower tier institutions, or have lower H-index scores (Lungeanu et al., 2014). Research also shows that prior collaboration reduces uncertainty about the likelihood that the pair will engage in a future collaboration (Cummings & Kiesler, 2008; Gruenfeld, Mannix, Williams, & Neale, 1996; Guimera, Uzzi, Spiro, & Amaral, 2005; Hinds, Carley, Krackhardt, & Wholey, 2000; Lungeanu & Contractor, 2015; Lungeanu et al., 2014). Finally, the patterns of relationships within the broader collaboration networks also affect the likelihood of collaboration between pairs of scientists. For example, Newman (2001, p. 408) showed that "friend-of-a-friend" mechanism (Heider, 1958) predicts future collaboration, with researchers having a "30% or greater probability of collaborating if both have collaborated with a third scientist."

Whereas prior work has built on the notion that collaboration occurs at the level of a *dyad*, in reality, collaboration takes place within *multilevel* social structures, with individuals engaged in *multiple teams* that each often have many more than two members (Bordons & Gómez, 2000). In fact, evidence suggests 65–90% of knowledge workers are members of multiple teams at any given time (O'Leary, Mortensen, & Woolley, 2011). Figure 1 provides an illustration of this "multiple team membership" phenomenon. This figure depicts ten individual scientists (labeled m1–m10 in orange) who are organized into six teams (labeled T1–T6 with lines indicating their boundary).

In Figure 1, let us assume two researchers (m5 and m6) have assembled into a newly formed team (Team 1), with boundaries indicated by the dotted red line. The figure indicates that in addition to Team 1, Member 5 also belongs to Team 2, and Member 6 belongs to both Teams 4 and 6. Thus, Team 1 is *directly* interlocked with Teams 2, 4, and 6 based on common members (i.e., a *member interlock*). In turn, the three teams in Team 1's *proximal* ecosystem (i.e., the team's with which it has direct member interlocks) include additional members who belong to other teams, and so on. Thus, as Figure 1 depicts, more distally, Team 1 is *indirectly* linked to Teams 3 and 5 through interlocks with teams that are directly interlocked with Team 1. In scientific ecosystems organized into teams,

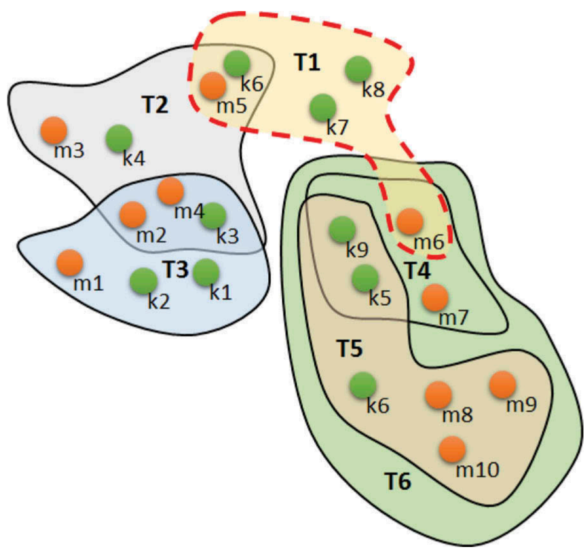


Figure 1. Sample representation of a scientific ecosystem characterized by interlocking teams. Note. The red dotted line represents the external boundary of a newly assembled scientific team; the solid black lines represent the boundaries of other scientific teams in the proximal and more distal surrounding ecosystem; A letter T represents a scientific team; A letter m represents a member of one or more scientific team; A letter k represents the knowledge domain considered within a scientific team.

knowledge "flows through network ties via the individuals that connect different teams by virtue of co-memberships on teams" (Zaheer & Soda, 2009, p. 3). Thus, the structures of team interlocks surrounding scientists are likely to be relevant to understanding the situations that give rise to new teams.

Figure 1 also depicts the knowledge domains (e.g., topics, areas of inquiry) relevant to each team (labeled k1–k9 in green), and the ways in which the teams are interlocked based on overlapping knowledge. Within the context of Team 1, Members 5 and 6 pursue research related to Knowledge Domains 6, 7, and 8. However, both scientists have also worked as members of teams that have considered other knowledge domains. As a member of Team 2, Member 5 investigates Knowledge Domains 3, 4, and 6. Thus, Team 1 is not only interlocked with Team 2 based on overlapping membership, the teams are joined through a *knowledge interlock* based on the overlapping engagement with Knowledge Domain 6. In fact, Member 5 might be bringing her prior experiences investigating Knowledge Domain 6 to bear within the context of Team 1. In contrast to Member 5, the teams that Member 6 has contributed to previously did not investigate any of the topics that are considered within Team 1. However, this scientist has had experience working within other knowledge domains that might be relevant to the topics of investigation within Team 1. Thus, the ways in which teams are interlocked through knowledge are relevant to understanding team assembly.

## *A hypergraph approach to characterizing team interlock ecosystems*

The team interlock view of scientific ecosystems, illustrated in Figure 1, depicts more accurately and richly the embedded social and intellectual milieu within which scientific collaboration occurs. However, the structure of team interlock ecosystems can be challenging to identify and describe. In fact, one reason for the paucity of studies examining overlapping teams is that such ecosystem-focused investigations entail multiple levels of analysis (e.g., individuals nested in multiple interlocking teams), which, in turn, requires the use of more complex statistical analyses and tools (Lungeanu, Sullivan, Wilensky, & Contractor, 2015). Network researchers interested in collectives have typically employed one of two analytic approaches to capture relational properties unfolding at multiple levels (e.g., person-to-person; person-to-team) of social systems. The first is to capture actor-to-actor relationships and represent their structure in a one-mode network; the second is to capture the linkages of actors-to-collectives in a two-mode or bipartite network linking individuals to teams. The former fails to represent the entitativity of the collective; put simply, three links between three nodes could imply one entity (or team) of three researchers or three entities (or teams) of three separate pairs of researchers. The latter, bimodal network approach, links individuals to collectives, but fails to capture individuals' relations with one another or other relations (e.g., overlap in knowledge domains) among teams.

In our conceptualization of collaboration, scientific teamwork constitutes more than a collection of dyadic person-to-person or person-to-team connections. Collaboration draws together a team of authors, publishing articles involving multiple knowledge areas. This collaboration cannot be accurately represented by projecting onto a one-mode network (researcher-to-researcher or person-to-person) or a two-mode network (researcher-to-article or person-to-team). Instead, the collaboration is better formalized as a ***hypergraph***, in which authors, keywords, and/or journals are combined in (possibly overlapping) sets (Shi, Foster, & Evans, 2015). Hypergraphs have been well established in the area of mathematics, as an extension to graph theory, beginning with foundational work by Berge (1973). Just as edges represent links between pairs of nodes within a network (or, what graph theorists call, a graph), ***hyperedges*** represent "links" connecting multiple (not necessarily a pair of) nodes that represent a single entity within a "hypergraph." For instance, in the case of a publication, a single hyperedge connects all of the authors on the team as well as all of the keywords for the article. A collection of hyperedges (with possible overlaps) constitutes a hypergraph, or what we refer to as a *team interlock ecosystem*. Recently, researchers have begun to recognize the value of hypergraphs as a means of representing and analyzing more complex data about teams, with promising results (Ghasemian, Zamanifar, & Ghasem-Aghaee, 2017; Ghasemian, Zamanifar, Ghasem-Aqaee, & Contractor, 2016; Sharma, Srivastava, & Chandra, 2014; Shi et al., 2015; Taramasco, Cointet, & Roth, 2010).

### Key components of hypergraphs

Hypergraph approaches are beneficial for characterizing the structures of team-based enterprises like scientific research. Here, we review key hypergraph components that can be used to characterize team interlock ecosystem structures.

First, in the context of scientific collaboration, a ***hyperedge*** represents the boundary of a scientific team such as those indicated by lines in Figure 1. Hyperedges are comprised of multiple ***nodes*** of one or more "type." For example, nodes could be researchers participating in a collaboration and the knowledge domains associated with that collaboration). Nodes can have connections to other nodes (as in one-mode person-to-person networks) as well as to hyperedges (as in bipartite person-to-team networks).

As a whole, a ***hypergraph*** constitutes a set of hyperedges which can be connected to one another based on node overlaps (i.e., member or knowledge interlocks). For example, Figure 1 depicts a hypergraph based on the publications involving members of a focal team. Mathematically, a hypergraph is represented as $H = (V, E)$. $V$ is a set of *nodes* (or vertices), $V = \{v_1, v_2, \ldots, v_n\}$, that can be authors, keywords, methods, etc. $E$ is a set of *hyperedges*, $E = \{e_1,\ e_2, \ldots, e_m\}$, that include as many or as few as zero nodes. This is an important departure from one-mode social networks in which edges are required to have exactly *two* nodes.

The interlocks between hyperedges based on overlapping nodes constitute ***hyperties***. A hypertie indexes the set of nodes shared by two or more hyperedges, which can include any arbitrary number of nodes in theory. Mathematically, we have: $T = \ e_1 \cap e_2 \cap \ \ldots \ \cap e_i$, for any $i$ number of hyperedges $e_1$ through $e_i$, where $i \geq 2$.

The ***local neighborhood*** $L(e)$ of a hyperedge $e$ is defined as $L(e) = \{h \in E : \exists v, s.t.\ v \subset \{e, h\}\}$, that is the set of all hyperedges $h$ (i.e., other teams) such that there is at least one node $v$ shared by both $e$ and $h$. In the case of scientific collaboration, the local neighborhood of a team is composed of the set of other teams that are directly connected to the focal team through at least one member or knowledge interlock.

### Key hypergraph characteristics

The core components of hypergraphs—nodes, ties, hyperedges, hyperties, and local neighborhoods—give rise to structural characteristics at multiple levels (e.g., node, hyperedge, hypergraph levels) (Sullivan, Zhu, Lungeanu, & Contractor, 2012). Thus, hypergraph approaches are particularly useful for characterizing the multilevel ecosystems with connections among people, knowledge, and/or other types of nodes organized into teams. Below we present a short description of these metrics. The full description of these metrics, together with the mathematical equations and the graphical representation, is available in the *Supplemental Material* Appendix.

#### Node metrics

At the nodal (i.e., person, keyword) level, we compute *a* ***node's degree*** and a ***node's hyperdegree***. *Node degree* refers to the number of distinct nodes with which a focal node is connected (Ghoshal, Zlatic, Caldarelli, & Newman, 2009; Wang, Rong, Deng, & Zhang, 2010), a *node's hyperdegree* is the number of hyperedges (e.g., teams) in which the node participates (Estrada & Rodríguez-Velázquez, 2006; Wang et al., 2010; Zlatic, Ghoshal, & Caldarelli, 2009).

#### Hyperedge metrics

At the level of a single team (i.e., a hyperedge), ***hyperedge degree*** reflects to the number of other hyperedges (teams) with the focal hyperedge (team) is interlocked via overlapping members

(Wang et al., 2010; Zlatic et al., 2009). Next, a ***hyperedge clustering coefficient*** reflects the degree to which the set of teams that are directly connected to the focal team—the team's "local" neighborhood—exhibit the network property of *triadic closure* among teams through interlock connections. A hyperedge clustering coefficient for a focal team represents the degree to which the set of teams that are directly interlocked with the focal team are, themselves, interlocked with one another.

### *Hypergraph metrics*

Hypergraph metrics move beyond characterizing patterns of connections surrounding individual nodes or individual teams to characterize the patterns of team interlocks for an entire scientific ecosystem (i.e., the hypergraph level). We consider three hypergraph metrics in particular: First, ***hypergraph density*** represents the proportion of interlocks among teams out of the total possible number of team interlocks (e.g., through overlapping membership; through overlapping knowledge). Second, a "***hyper*graph *clustering coefficient***" which indicates the degree to which *all* possible triads of teams in a hypergraph exhibit the property of closure. A final hypergraph metric, ***hypergraph centralization*** indicates the degree of variance in the distribution of hyperedge degree centrality scores across a hypergraph.

## Illustrative hypotheses linking characteristics of team interlock ecosystems to scientific team assembly

Adopting a team-interlock perspective to conceptualize scientific collaboration and a hypergraph methodological approach to characterize structures of scientific ecosystems enables researchers to more accurately understand the ecosystem factors influencing team assembly. Research on teams has shown that important phenomena related to social bonding (e.g., cohesion, trust), knowledge generation, and learning emerge at the *level of the team* as a whole, and that teams often differ from one another substantially in terms of their team-level properties (Edmondson, 1999; Kozlowski & Klein, 2000). Thus, we propose that in addition to person-to-person connections among researchers, the patterns of interlock connections between teams are relevant to team assembly. The overarching question addressed by this research is: *What structural characteristics of scientific ecosystems affect the likelihood that sets of researchers will assemble into a new team*?

To demonstrate how this research question might be addressed, we provide illustrative hypotheses considering the extent to which team assembly is influenced by three of the key structural characteristics of scientific ecosystems: (1) *hypergraph clustering coefficient; (2) hyperedge clustering coefficient;* and (3) *hypergraph centralization*. These three characteristics are similar to three characteristics of person-to-person social networks (e.g., Oh, Chung, & Labianca, 2004; Oh, Labianca, & Chung, 2006) that have been shown to have important implications for individual and collective outcomes (i.e., "social network clustering", "individual brokerage," and "social network centralization," respectively). We extend work on person-to-person social networks to explore three research questions considering the degree to which the presence of these properties within scientific ecosystems influences team assembly.

## Research Question 1: Are researchers more likely to assemble into a team in ecosystems characterized by greater "coherence" (i.e., higher hypergraph clustering coefficient scores)?

Substantively, a hypergraph clustering coefficient characterizes the extent to which teams tend to be organized into ***coherent*** intellectual communities, whose membership and knowledge overlap. Ecosystem coherence has the potential of impacting the assembly of new scientific teams because the creation of new knowledge is a result of a social process in which individual researchers share expertise and gain legitimacy by working across overlapping teams (Acedo, Barroso, Casanueva, & Galán, 2006; Moody, 2004; Wuchty et al., 2007).

The process of sharing and learning diverse knowledge, however, can be complicated by the loss of meaning during the transfer process (Grant, 1996). As a result, scholars argue that the transfer and combination of knowledge needs to be facilitated by the development of a common language and exchange norms (Grant, 1996; West & Anderson, 1996). A more coherent ecosystem structure—one with higher levels of clustering—reflects the presence of a community of scientists who have built common socio-cognitive models that allow them to adopt a common language for the knowledge that they exchange. Such cognitive models are necessary in order to recognize the knowledge held by others, understand current knowledge sharing practices, and understand the rules to identify new and useful knowledge and recombine prior knowledge in order to generate new ideas (Murray & O'Mahony, 2007; Uzzi & Spiro, 2005). For example, Murray and O'Mahony (2007) argue that in order for innovation to occur, existing knowledge must be shared within intellectual communities, but, in addition, knowledge must be reused, recombined, and accumulated. Higher levels of ecosystem coherence offer opportunities for scientists to share, search for, access, and apply knowledge. Accordingly, we hypothesize that scientific teams are more likely to assemble in ecosystems that are coherent by virtue of having triadic closure among teams based on overlapping members, overlapping knowledge domains or both (the intersection of the team and knowledge interlocks).

*Hypothesis 1: Scientific teams are more likely to assemble in ecosystems that are coherent by virtue of having triadic closure among teams based on overlapping members.*

*Hypothesis 2: Scientific teams are more likely to assemble in ecosystems that are coherent by virtue of having triadic closure among teams based on overlapping knowledge domains.*

*Hypothesis 3: Scientific teams are more likely to assemble in ecosystems that are coherent by virtue of having triadic closure among teams based on having both overlapping members and overlapping knowledge domains.*

In summary, we have proposed three hypotheses linking scientific ecosystems to the assembly of teams. The first specified the effect of ecosystem coherence in team member interlocks, whereas the second specified the effect of ecosystem coherence in knowledge interlocks. Understanding the joint effects of ecosystem coherence captured by the intersection of member and knowledge interlocks (Hypothesis 3) is important because it prompts us to further inquire: *Which of these structural properties is more important for team assembly: ecosystem coherence with regard to: (a) team member interlocks, (b) team knowledge interlocks, or (c) intersection of team member and knowledge interlocks?*

Next, we turn to the second of our three research questions exploring how ecosystems influence team assembly.

## Research Question 2: Are researchers more likely to assemble into a team when their local neighborhoods exhibit greater "local brokerage" (i.e., lower hyperedge clustering coefficient scores)?

Although our first three hypotheses posit that closure across the *broader* ecosystem (i.e., ecosystem coherence) support team assembly, scholars have also pointed out that excessive closure might hurt creativity. In fact, research suggests that having a sufficient level of diversity in knowledge and social connections is a critical factor underpinning creative and/or innovate ideas (Fleming, Mingo, & Chen, 2007; Guimera et al., 2005; Reagans, Zuckerman, & McEvily, 2004; Uzzi & Spiro, 2005). Substantively, a hyperedge clustering coefficient metric represents the degree of diversity (or lack thereof) in team memberships and expertise in a team's local community. When there are no interlocks between the set of teams that are interlocked with a focal team, the hyperedge clustering coefficient for that focal team would be zero. In this case, the members of the focal team are uniquely positioned to draw upon and combine *different* resources from other teams of which they are also members. Thus, a lower hyperedge clustering coefficient is an indicator that the team is a broker between other teams.

In person-to-person networks, research suggests individuals who occupy positions of brokerage —those with connections to people who themselves are unconnected—can reap career benefits, in part, because they have greater access to diverse ideas (Burt, 1992). Likewise, we expect that a team of scientists whose local ecosystem communities exhibit *lower* levels of closure (i.e., higher brokerage at the team-level) are more likely to have access to diverse ideas that lead to new research projects. Hence, our next set of hypotheses posit teams will be more likely to assemble when there is sufficient opportunity for potential team members to broker ideas from other teams they belong to, but which do not have overlapping members (besides them) or overlapping knowledge domains. Accordingly:

*Hypothesis 4: Scientific teams are more likely to assemble when they are brokering ties within the local neighborhood of their team member interlock ecosystem.*

*Hypothesis 5: Scientific teams are more likely to assemble when they are brokering ties within the local neighborhood of their team knowledge interlock ecosystem.*

*Hypothesis 6: Scientific teams are more likely to assemble when they are brokering ties within the local neighborhood of their team member and knowledge interlock ecosystem.*

Whereas these hypotheses test specific predictions rooted in prior research on network brokerage, we also explore the question of which structural properties are more important to team assembly: *local brokerage with regard to: (a) team member interlocks, (b) team knowledge interlocks, or (c) team member and knowledge interlocks?*

Finally, we turn to the last of our three research questions exploring how ecosystems influence team assembly.

## Research Question 3: Are researchers more likely to assemble into a team in ecosystems characterized by greater "ecosystem decentralization" (i.e., lower hypergraph centralization scores)?

A final consideration when sets of scientists decide to assemble into a new team is whether there are sufficient opportunities to impact the broader scientific community through the generation of scientific output. As Murray and O'Mahony (2007) note, expectation of reward is a key consideration when engaging in creative work. One ecosystem characteristic that may impact whether a set of scientists expect rewards for assembling into a new team is the degree to which the ecosystem is ***centralized*** around one or a few teams that are interlocked with many other teams (i.e., hypergraph centralization).

On the one hand, well-developed theories and concepts are beneficial for the advancement of scientific fields because they provide a sense of direction for knowledge development and enable understanding of key research topics (Kuhn, 1996). High ecosystem centralization indicates that a few teams of productive individuals have successfully achieved the development of conceptual frameworks and have disseminated those ideas by collaborating with many other teams. As such, higher ecosystem centralization signify ecosystems with less opportunity for impact. Thus, scientists operating in highly centralized ecosystems may not have the motivation to assembly into a new team, unless the new team does contains members of a highly central team. Thus, we posit that when ecosystems are more ***de*-centralized** (i.e., less centralized), scientific teams are more likely to assemble.

*Hypothesis 7: Scientific teams are more likely to be assembled when the ecosystem formed by team member interlock is highly decentralized.*

*Hypothesis 8: Scientific teams are more likely to be assembled when the ecosystem formed by team knowledge interlock is highly decentralized.*

*Hypothesis 9: Scientific teams are more likely to be assembled when the ecosystem formed by team member and knowledge interlock is highly decentralized.*

Here, again, we explore the question of which of these effects is more important: *Decentralization in a (a) team member interlock, (b) team knowledge interlock, or (c) team member and knowledge interlock ecosystem?*

## Method

We deploy a novel hypergraph methodology to test our hypotheses using bibliographic data about teams submitting research proposals to a Clinical and Translational Science Award (CTSA) competition hosted at a large Midwestern University and funded by the National Institutes of Health (NIH). A total of 101 research teams, consisting of 147 participants, submitted proposals in two rounds of the grant competition. Given that we are examining the team assembly process, we excluded 47 proposals that were solo-authored. Additionally, eight proposals were excluded because either the exact same proposal team submitted proposals in both rounds of the competition (three teams) or because of data collection issues (five teams). The final dataset contains 46 proposals co-authored by 98 scientists, out of which only four proposals were awarded.

For each proposal team, we extracted team members' collaborators and the collaboration (co-authorship) relations among those collaborators. First, we used the *Web of Science (WoS)* database provided by Thomson Reuters to extract each team member's publication history. Author name disambiguation is a recognized issue when constructing bibliometric measures (Torvik, Weeber, Swanson, & Smalheiser, 2005). This is the problem of ensuring that we only consider, for instance, the publications by John Smith who submitted a research proposal and not others with the same name. To overcome this limitation, we manually verified each publication against researchers' resumes available on the institution website. Second, we identified all co-authors listed on the above publications and disambiguated their names. For example, "Smith, J" and "Smith, JH" were considered same person. This is because the probability that one researcher will collaborate with two different researchers named "Smith, J" and "Smith, JH" is very low. Third, we extracted all publications of the extended list of co-authors from the WoS database. We considered only those publications that were co-authored by at least two researchers from our dataset to be valid.

Based on this information, we created 46 unique team interlock ecosystems (i.e., one for each proposal team). Each interlock ecosystem included the focal proposal team, all of the focal team's interlocking teams (all publications co-authored by members of the focal team), and all of the second-order interlocking teams connected to those initial interlocking teams (all publications co-authored by co-authors of the focal team). Additionally, for each publication included in the ecosystem, we extracted both the "author keywords" (i.e., keywords provided by the original authors) and the "keywords plus" (i.e., keywords extracted from the titles of the cited references by Thomson Reuters) available in the WoS database. Therefore, each team ecosystem is formed from teams (i.e., hyperedges) that contain two type of nodes: scientists and keywords. The teams are linked together through a team membership interlock, team knowledge (keywords) interlock, or the intersection of the team membership and knowledge interlock. It is important to note that this depiction of the scientific ecosystem begins with interlocks based on overlapping team membership and then measures the degree to which teams with interlocking membership also have overlap on the topics studied by the focal team.

Figures 2a–2c present the visual representation of one of the 46 team interlock ecosystems from our dataset. As shown in this figure, the team interlocks (i.e., hyperties between teams) differ when we consider a network based on member interlock, knowledge interlock, or the intersection of the member and knowledge interlock.

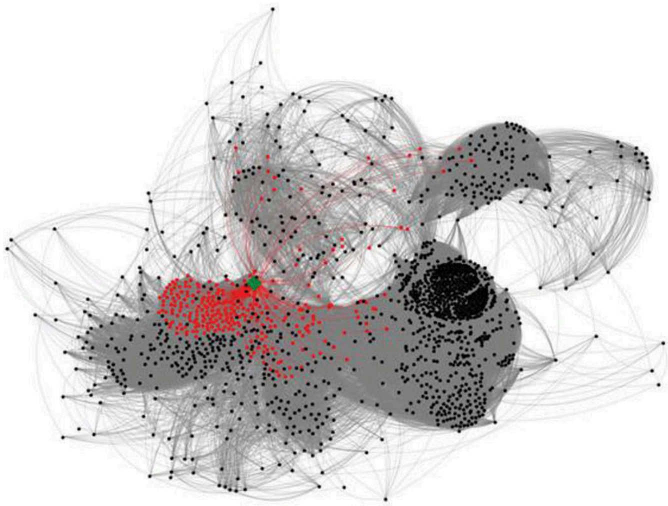

**Figure 2a.** A random sample team ecosystem from our dataset (ecosystem with 1612 teams, 246 scientists, and 5,462 keywords). The green diamond node represents the focal team (team that was assembled). The red dots represent teams in the local ecosystem (with which they had direct overlapping members). The black dots represent the remaining teams in the ecosystem (with which they had indirect overlapping members via the teams represented by the red dots). The links are based on team member interlock.

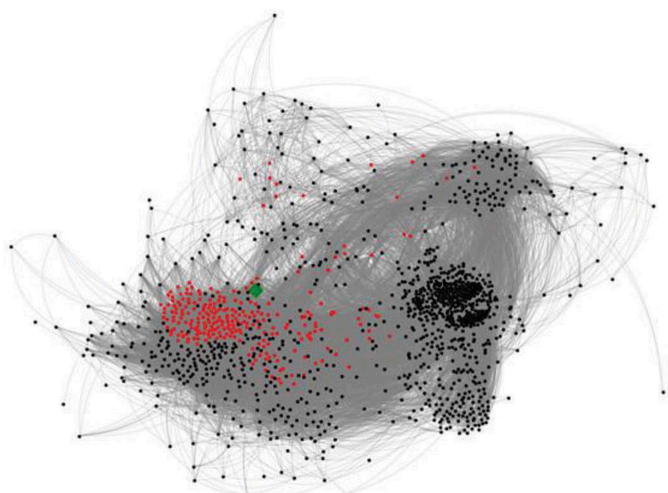

**Figure 2b.** A random sample team ecosystem from our dataset (ecosystem with 1612 teams, 246 scientists, and 5,462 keywords). The green diamond node represents the focal team (team that was assembled). The red dots represent teams in the local ecosystem (with which they had direct overlapping members). The black dots represent the remaining teams in the ecosystem (with which they had indirect overlapping members via the teams represented by the red dots). The links are based on team knowledge interlock.

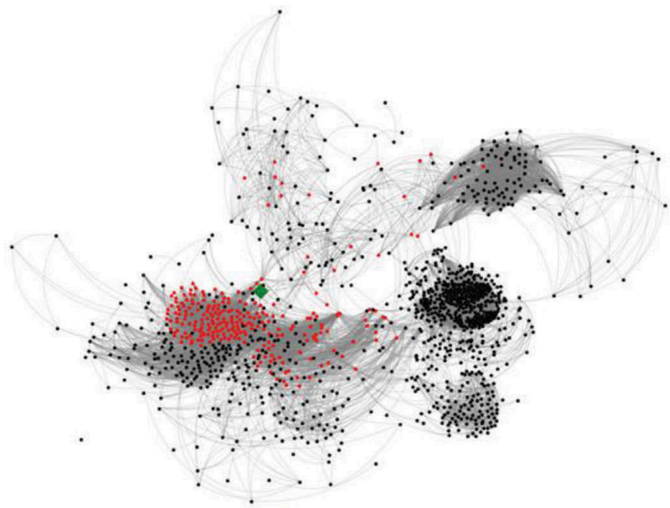

Figure 2c. A random sample team ecosystem from our dataset (ecosystem with 1612 teams, 246 scientists, and 5,462 keywords). The green diamond node represents the focal team (team that was assembled). The red dots represent teams in the local ecosystem (with which they had direct overlapping members). The black dots represent the remaining teams in the ecosystem (with which they had indirect overlapping members via the teams represented by the red dots). The links are based on team member and knowledge interlock both being present.

## *Hypergraph indices representing team interlock ecosystem characteristics*

In order to characterize properties of coherence, brokerage, and centralization for each of the 46 team interlock ecosystems, we computed a set of hypergraph-based descriptive metrics corresponding to each concept. Ecosystem coherence was computed using the *hypergraph clustering coefficient* measure. As mentioned previously, hypergraph clustering coefficient indicates the degree to which all possible triads of teams exhibit closure across a hypergraph. A high hypergraph clustering coefficient indicates that there is a high amount of overlap or "coherence" across the ecosystem, based on teams that share common members (team member interlock), common knowledge (team knowledge interlock), or the combination of members and knowledge (team member and knowledge interlock).

Local brokerage was computed using the *hyperedge clustering coefficient* metric. The hyperedge clustering coefficient is defined as the amount of overlap that exists among the teams that share common members with the focal team (team member interlock), or amount of overlap that exists among the teams that share the same keywords (team knowledge interlock) or both (the intersection of the team member and knowledge interlock). It is computed as the density of ties among the alters of the focal team. For instance, when there is no overlapping team membership among the teams with which the focal team members co-authored, the hyperedge clustering coefficient would be zero. In other words, high clustering (or low brokerage) means that researchers tend to collaborate with the collaborators of their collaborators.

Ecosystem decentralization was computed using the *hypergraph centralization* measure. Hypergraph centralization indicates the degree of variance in the distribution of teams' degree centrality scores across the ecosystem. In other words, a high ecosystem centralization implies that one or a few teams have a disproportionate number of team interlock connections with other teams based on overlapping membership and/or overlapping knowledge domain.

## *Ecosystem simulation analytic approach*

The aforementioned descriptive metrics were computed for each of the 46 ecosystems in which each of the 46 focal (proposal) teams were embedded. However, by themselves these descriptive measures do not tell us if the relevant metric is high or low—as compared to chance, captured by a null model. Therefore, to test our hypotheses, we developed a computational technique to compare the observed team ecosystems with simulated *null models*. For each team ecosystem, which is a network of teams, we began by generating 200 simulated synthetic networks which served as the null models. The output of the simulation included means, $\delta_{sim}$, and standard deviations, $\sigma(\delta_{sim})$, for the following measures: hypergraph clustering coefficient, hyperedge clustering coefficient of the focal team, and hypergraph centralization. Second, we tested whether the frequency distribution of these measures in the null models were normally distributed. Third, we tested our hypotheses by comparing the ecosystem metrics generated from the simulated networks with the observed ecosystem. Specifically, we computed the z-scores of the observed measures relative to the random measures: $z(\delta_{obs}) = (\delta_{obs} - \delta_{sim})/\sigma(\delta_{sim})$.

### Null model simulation

For each of the 46 focal team interlock ecosystems, we created a null model reflecting a set of random synthetic networks that incorporate realistic aspects of the observed data and its network structure. In particular, the null model for each ecosystem was based on generating synthetic networks that shared the following empirical facts with the observed ecosystems: the number of teams, the number of authors per team, and the number of keywords per team. Therefore, the null model preserves for each ecosystem the same number of teams, authors, and keywords as our observed ecosystems.

In order to test the team member interlock hypotheses (H1, H4, and H7), we started from the observed network and generated 200 random samples, while holding constant the number of authors, the number of publications, the distribution number of authors per publication, and the distribution of number of publications per author. In order to test the team knowledge interlock hypotheses (H2, H5, and H8), we generated 200 random samples, fixing the number of keywords, the number of publications, the distribution number of keywords per publication, and the distribution of number of publications per keyword. Finally, in order to test the team member and knowledge interlock hypotheses (H3, H6, and H9), we generated 200 random samples, fixing the number of authors and keywords, the number of publications, the distribution number of authors and keywords per publication, and the distribution of number of publications per author and per keyword.

For each of the 200 simulated ecosystems, we computed the same hypergraph metrics we obtained for the observed ecosystem. We then computed a z-score comparing each of the observed hypergraph metrics with the distribution of that corresponding metric in the 200 simulated ecosystems. A positive z-score indicates a score that is higher than expected by chance. A negative z-score indicates a score that is less than expected by chance. The larger the z-score, the greater the difference between the observed score and the mean score obtained from simulations. Since utilizing the z-score assumes that the data is a normally distributed, we tested the distribution of each of the hypergraph metrics for each ecosystem. We computed Skewness/Kurtosis values for each metric obtained from the simulation of the 46 team ecosystems. Using Bulmer's (1979) rule of thumb, we assessed if each of the distributions were approximately symmetric (i.e., skewness is between −½ and +½,) or moderately skewed (i.e., skewness is between −1 and −½ or between +½ and +1) for all ecosystems. Furthermore, we also conducted the Kolmogorov-Smirnov test for normality which yields a p-value for each metric for each ecosystem. A p-value higher than 0.05 implies that the distribution is normal. When a distribution is deemed not to be normal, we do not report the z-score results.

## Results

### Descriptive statistics

Table 1 and Figures 3–5 provide descriptive information about our team ecosystems. We examine traditional descriptive statistics such as the number of observed teams, authors, and keywords, and features of team interlocks. We also examine whether the team interlock structures follow a power law distribution.

**Table 1.** Descriptive statistics.

| Variable | Metric | Mean | SD | MIN | MAX |
|---|---|---|---|---|---|
| Proposal team: Team size | | 2.39 | 0.61 | 2 | 4 |
| **Ecosystem - general measures** | **Hypergraph - general measures** | | | | |
| Ecosystem: Number of teams in the ecosystem | Number of hyperedges in the hypergraph | 2441.93 | 2572.52 | 47 | 8576 |
| Ecosystem: Number of authors in the ecosystem | Number of nodes (node type = scientist) in the hypergraph | 242.67 | 205.02 | 14 | 925 |
| Ecosystem: Number of keywords in the ecosystem | Number of nodes (node type = keyword) in the hypergraph | 5420.13 | 4720.11 | 238 | 15589 |
| Local ecosystem: Number of teams (direct links) to the focal team | Local neighborhood: number of hyperedges with direct hyperties to the focal hyperedge | 143.78 | 122.81 | 12.00 | 615.00 |
| Local ecosystem: Number of unique teams (direct links) to the focal team | Local neighborhood: number of unique hyperedges with direct hyperties to the focal hyperedge | 140.28 | 119.96 | 12.00 | 596.00 |
| **Team member interlock** | | | | | |
| Density | Hypergraph Density | 0.15 | 0.11 | 0.02 | 0.44 |
| Centralization | Hypergraph Centralization | 0.26 | 0.09 | 0.11 | 0.43 |
| Ecosystem coherence | Hypergraph Clustering coefficient | 0.85 | 0.05 | 0.74 | 0.95 |
| Local closure | Hyperedge (local) clustering coefficient | 0.61 | 0.15 | 0.32 | 0.93 |
| **Team knowledge interlock** | | | | | |
| Density | Hypergraph Density | 0.05 | 0.07 | 0.01 | 0.35 |
| Centralization | Hypergraph Centralization | 0.15 | 0.07 | 0.07 | 0.33 |
| Ecosystem coherence | Hypergraph Clustering coefficient | 0.57 | 0.05 | 0.50 | 0.75 |
| Local closure | Hyperedge (local) clustering coefficient | 0.16 | 0.15 | 0.02 | 0.72 |
| **Team member & knowledge interlock** | | | | | |
| Density | Hypergraph Density | 0.02 | 0.04 | 0.00 | 0.18 |
| Centralization | Hypergraph Centralization | 0.09 | 0.07 | 0.01 | 0.28 |
| Ecosystem coherence | Hypergraph Clustering coefficient | 0.60 | 0.07 | 0.40 | 0.76 |
| Local closure | Hyperedge (local) clustering coefficient | 0.11 | 0.11 | 0.01 | 0.50 |

N = 46 team ecosystems

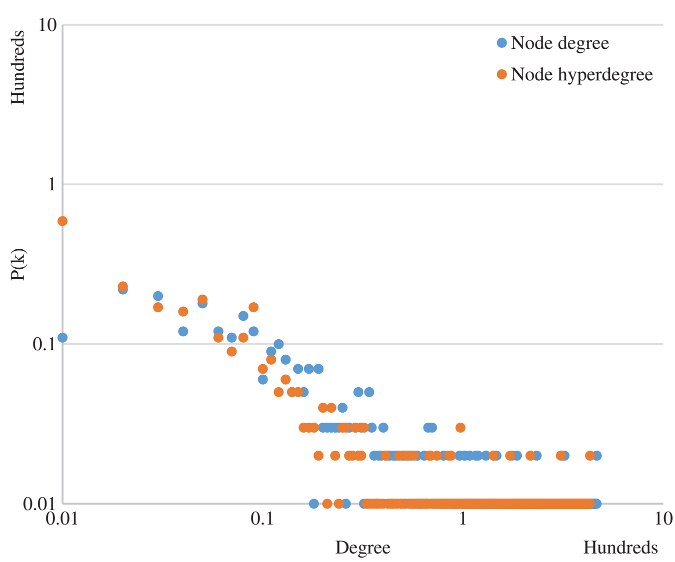


**Figure 3.** Degree distribution for node type scientist (team ecosystem T024).

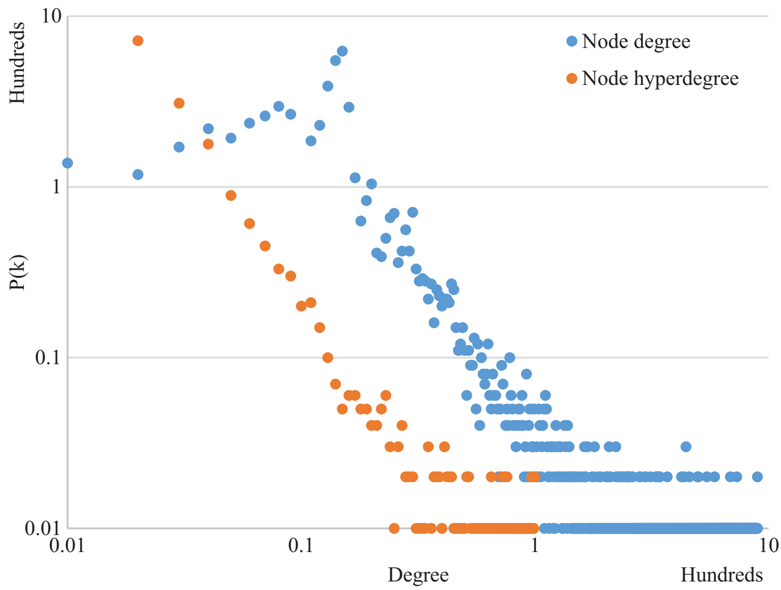


**Figure 4.** Degree distribution for node type keyword (team ecosystem T024).

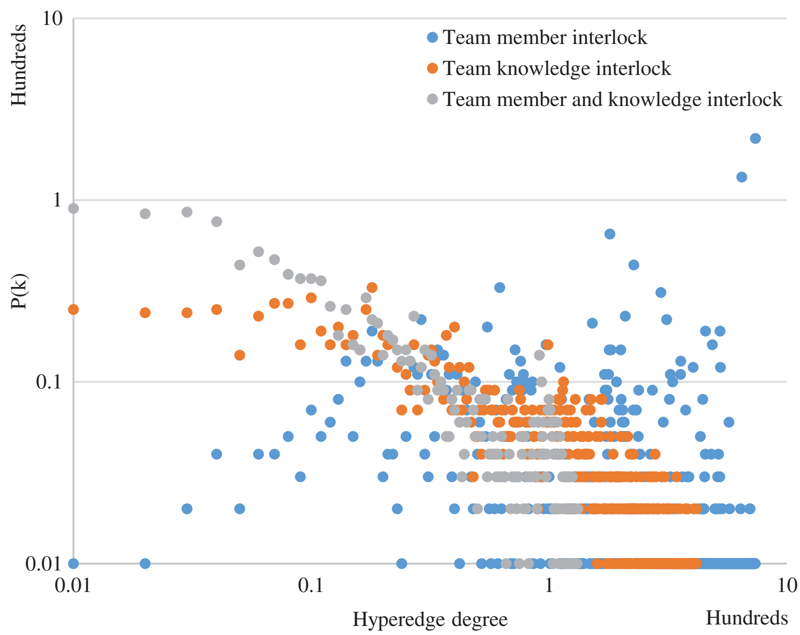


**Figure 5.** Hyperedge degree distribution (team ecosystem T024).

## *Team descriptive statistics*

The team ecosystems ranged from being comprised of between 47 and 8,576 teams (M = 2,441.93, SD = 572.52), each with between 14 and 925 unique authors (M = 242.67, SD = 205.02), and between 238 and 15,589 unique keywords (M = 5,420.13, SD = 4,720.11). Table 1 presents the overall descriptive statistics for the observed team ecosystems we created around each focal proposal.

Generally, all hypergraph metrics decrease as we move from team member interlocks to team knowledge interlocks to the combined team member and knowledge interlocks. The ecosystems have low density, and, as expected, the density is lowest for the combined team member and knowledge interlocks, because there are fewer hyperties between teams when the ecosystems are based on the presence of both, that is the intersection of, member *and* knowledge interlocks. Furthermore, the ecosystems have an average clustering coefficient of 0.85 (team member interlocks), 0.57 (team knowledge interlocks), and, respectively, 0.60 (combination of team member and knowledge interlocks).

As discussed previously, a hypergraph approach enables us to compute different degree metrics: node degree, node hyperdegree, and hyperedge degree. To illustrate the differences between node degree and node hyperdegree we use one randomly selected team ecosystem from our dataset. A quick examination of these measures provides interesting insights. For example, in the ecosystem for team T024, a scientist can be linked to a maximum of 466 scientist (i.e., maximum node degree for node type scientist in ecosystem T024) and to a maximum of 433 teams (i.e., maximum node hyperdegree for node type scientist in ecosystem T024). See Figure 3 for a presentation of the degree distribution for node type scientist. However, a keyword can be linked to maximum of 998 other keywords (i.e., maximum node degree) but only to a maximum 100 teams (i.e., maximum node hyperdegree). Figure 4 presents the degree distribution for node type keywords. This is an important distinction when we discuss member or knowledge overlap and how far knowledge can spread. This property of knowledge spread relative to team member spread raises an interesting feature of ecosystems. Whereas members have fewer other members to reach out to, they have far more knowledge domains with which they can connect.

## Power law distribution

Next, we examined how team interlocks are distributed in the ecosystem. Specifically, we explored whether the team interlock structures exhibit a specific pattern, such as a power law distribution (Barabási & Albert, 1999), i.e., $P(k) \sim k^{-\gamma}$, where $k$ is the degree and $P(k)$ is the fraction of hyperedges that have the degree $k$. A power law distribution will show that the most hyperedges (i.e., teams) have a low degree and a few hyperedges have a very high degree. This situation will indicate a hierarchy in the ecosystem. Figure 5 presents the hyperedge degree distribution. None of the team interlock ecosystems in our sample followed a power law distribution. This could have been an artifact of the way we constructed the teams' ecosystems: We started from the focal team and considered only those teams (i.e., scientific articles), authors and keywords, that are linked to the focal team directly in one step or indirectly in two steps. Therefore, we excluded the collaborating teams for those teams that are more than two-steps away from the focal team and hence do not belong to the team ecosystem.

## Inferential statistics: Hypothesis testing

Before interpreting the results of our hypotheses, it is important to acknowledge that our analysis examines the ecosystem structures of those teams that successfully assembled and submitted grant proposals. Therefore, our analysis does not include those researchers who intended to collaborate and submit proposals, but never submitted the proposals. However, given that we include teams that submitted proposals in two rounds of the grant competition, we are cautiously confident that researchers who had the intention to submit a proposal and had started to collaborate on writing the proposal, had the opportunity to submit the grant proposal in the specified time frame. Furthermore, as discussed above, to partially address this limitation, our analysis compares the observed ecosystem of the assembled team with a set of random simulated ecosystems that match basic characteristics of the observed ecosystem in terms of members, teams, and knowledge areas.

### *Ecosystem coherence*

Our first set of analyses examine the level of coherence in the scientific ecosystem, and the overarching Research Question 1: Are scientists more likely to assemble into a team in ecosystems characterized by higher levels of clustering (i.e., greater ecosystem coherence)? Table 2 presents the hypergraph clustering coefficient for the observed ecosystems and their z-scores based on 200 simulations. A positive z-score indicates a score that is higher than expected by chance. A negative z-score indicates a score that is less than expected by chance. The larger the z-score, the greater difference there is between the score and the mean score obtained from simulations.

Hypothesis 1 posited that scientific teams are more likely to assemble within team ecosystems characterized by greater coherence in team member interlocks. The simulation results showed that the observed hypergraph clustering coefficient in the team member interlock network is higher than expected by chance for all 46 team ecosystems. The results support Hypothesis 1: team ecosystem coherence increases the likelihood for a team to assemble.

Hypothesis 2 posited that scientific teams are more likely to assemble within team ecosystems characterized by greater coherence in team knowledge interlocks. The simulation results showed that the observed hypergraph clustering coefficient in the team knowledge interlock network is higher than expected by chance for 21 team ecosystems and lower than expected by chance for 25 team ecosystems. Interestingly, our hypothesis is supported only for small size team ecosystems. For large ecosystems, ecosystem coherence in team knowledge interlocks does not predict team assembly. This result is explained by the fact that large ecosystems contain interdisciplinary teams that are composed of researchers from different disciplines that publish both single discipline and interdisciplinary articles. The interdisciplinary articles are linked to multiple single discipline articles by virtue of using the same keywords, but the single discipline articles are not interlocked. Therefore, the coherence based on knowledge interlock is lower than expected by chance in such large ecosystems.

Hypothesis 3 posited that scientific teams are more likely to assemble within team ecosystems characterized by greater coherence in the intersection of the team member *and* knowledge interlocks. The simulation results showed that the observed hypergraph clustering coefficient in the member and knowledge interlock network is higher than expected by chance for all 46 team ecosystems. The results support our hypothesis: team ecosystem coherence, where the interlock represents the intersection of the member and knowledge interlock network increases the likelihood for a team to assemble.

We concluded our investigation of the effect of ecosystem coherence on team assembly with the exploratory research question: Which interlock type is more important? To address this question, we conducted paired sample t-tests to compare each pair of ecosystem clustering coefficient scores: team member interlocks, team knowledge interlocks, and the intersection of member and knowledge interlocks. Examining the results of these t-tests shows the hypergraph clustering coefficient for team member interlock networks ($M = 0.85$, $SD = 0.01$) is significantly higher than the hypergraph clustering coefficient effect for team knowledge interlock networks ($M = 0.57$, $SD = 0.01$); $t(45) = 31.92$, $p = 0.000$, and significantly higher than the hypergraph clustering coefficient for the intersection of the team member and knowledge interlock networks ($M = 0.60$, $SD = 0.01$); $t(45) = 30.17$, $p = 0.000$. Furthermore, paired t-test results show that the hypergraph clustering coefficient for team knowledge interlock networks ($M = 0.57$, $SD = 0.01$) is significantly lower than the hypergraph clustering coefficient for the intersection of the team member and knowledge interlock networks ($M = 0.60$, $SD = 0.01$); $t(45) = -2.99$, $p = 0.002$. These results suggest that high levels of ecosystem coherence with regard to team member interlock networks is most important to team assembly, and relatively more so than high levels of ecosystem coherence for the intersection of the member and knowledge interlock networks or for team knowledge interlock networks alone.

**Table 2.** Hypergraph clustering coefficient.

| | Team member interlock | | Team knowledge interlock | | Team member & knowledge interlock | |
|---|---|---|---|---|---|---|
| Team Id | Observed value | Z-score | Observed value | Z-score | Observed value | Z-score |
| T001 | 0.86 | 121.16*** | 0.75 | 32.38*** | 0.61 | 27.36*** |
| T002 | 0.93 | 228.69*** | 0.68 | 7.05*** | 0.76 | 105.56*** |
| T003 | 0.88 | 194.66*** | 0.59 | 44.67*** | 0.65 | 83.02*** |
| T004 | 0.85 | 280.19*** | 0.64 | 53.88*** | 0.73 | 157.05*** |
| T005 | 0.89 | 284.23*** | 0.65 | 75.99*** | 0.68 | 129.04*** |
| T006 | 0.91 | 223.14*** | 0.64 | 71.70*** | 0.67 | 72.02*** |
| T007 | 0.95 | 291.49*** | 0.69 | 139.24*** | 0.74 | 163.99*** |
| T008 | 0.93 | 812.33*** | 0.55 | 40.98*** | 0.56 | 125.09*** |
| T009 | 0.93 | 600.68*** | 0.64 | 29.04*** | 0.64 | 104.34*** |
| T010 | 0.88 | 352.07*** | 0.56 | 43.90*** | 0.61 | 192.64*** |
| T011 | 0.87 | 759.84*** | 0.58 | −9.90*** | 0.63 | 242.25*** |
| T012 | 0.80 | 423.79*** | 0.55 | 55.97*** | 0.61 | 233.71*** |
| T013 | 0.85 | 839.32*** | 0.59 | −50.24*** | 0.61 | 258.03*** |
| T014 | 0.93 | 1284.03*** | 0.50 | 41.77*** | 0.64 | 426.10*** |
| T015 | 0.86 | 830.83*** | 0.50 | −61.17*** | 0.57 | 202.08*** |
| T016 | 0.81 | 142.83*** | 0.60 | 30.23*** | 0.60 | 156.36*** |
| T017 | 0.89 | 354.23*** | 0.58 | 149.31*** | 0.66 | 318.89*** |
| T018 | 0.88 | 777.47*** | 0.52 | −10.65*** | 0.63 | 336.57*** |
| T019 | 0.85 | 709.02*** | 0.57 | 90.24*** | 0.62 | 314.02*** |
| T020 | 0.94 | 1124.12*** | 0.62 | −92.02*** | 0.69 | 315.51*** |
| T021 | 0.88 | 551.96*** | 0.51 | −34.63*** | 0.54 | 209.73*** |
| T022 | 0.81 | 507.98*** | 0.55 | 175.28*** | 0.43 | 140.96*** |
| T023 | 0.82 | 1013.03*** | 0.60 | −133.95*** | 0.59 | 252.39*** |
| T024 | 0.86 | 556.24*** | 0.60 | 206.00*** | 0.64 | 417.94*** |
| T025 | 0.89 | 630.19*** | 0.60 | 43.81*** | 0.67 | 312.78*** |
| T026 | 0.77 | 684.44*** | 0.55 | 120.19*** | 0.61 | 396.07*** |
| T027 | 0.83 | 947.78*** | 0.54 | −58.12*** | 0.55 | 312.86*** |
| T028 | 0.82 | 821.08*** | 0.60 | −49.84*** | 0.61 | 304.41*** |
| T029 | 0.88 | 713.93*** | 0.52 | −261.31*** | 0.55 | 228.93*** |
| T030 | 0.88 | 1317.14*** | 0.58 | −69.06*** | 0.66 | 409.78*** |
| T031 | 0.84 | 930.66*** | 0.54 | −24.71*** | 0.61 | 377.31*** |
| T032 | 0.86 | 876.08*** | 0.51 | −326.27*** | 0.56 | 234.78*** |
| T033 | 0.83 | 1197.30*** | 0.57 | −102.37*** | 0.59 | 555.33*** |
| T034 | 0.87 | 1325.66*** | 0.53 | −38.17*** | 0.63 | 491.63*** |
| T035 | 0.86 | 1565.94*** | 0.58 | −289.57*** | 0.65 | 517.23*** |
| T036 | 0.78 | 785.48*** | 0.55 | 25.58*** | 0.57 | 466.46*** |
| T037 | 0.86 | 1665.84*** | 0.55 | −298.17*** | 0.60 | 499.70*** |
| T038 | 0.79 | 483.72*** | 0.56 | 87.82*** | 0.53 | 506.50*** |
| T039 | 0.82 | 980.39*** | 0.56 | −201.52*** | 0.56 | 462.03*** |
| T040 | 0.78 | 1578.92*** | 0.54 | −449.68*** | 0.54 | 578.55*** |
| T041 | 0.79 | 1923.84*** | 0.57 | −165.54*** | 0.43 | 279.81*** |
| T042 | 0.82 | 1848.76*** | 0.54 | −409.16*** | 0.62 | 853.42*** |
| T043 | 0.79 | 2045.62*** | 0.57 | −476.72*** | 0.58 | 675.79*** |
| T044 | 0.74 | 1956.96*** | 0.55 | −900.15*** | 0.53 | 544.91*** |
| T045 | 0.78 | 2031.74*** | 0.55 | −645.67*** | 0.40 | 233.40*** |
| T046 | 0.79 | 2090.51*** | 0.52 | −579.41*** | 0.56 | 702.76*** |

*Notes:*
1. Teams are sorted based on hypergraph size (from small to large).
2. Z-score is based on 200 simulations. Z-score is presented only for normal distribution frequencies. + 0.10 * 0.05 ** 0.01 *** 0.001.

## Local brokerage

Our second set of analyses examined the level of local brokerage and the overarching Research Question 2: Are scientists more likely to assemble into a team when their local ecosystems are characterized by lower levels of clustering (i.e., greater local brokerage)? For these analyses, we compared the local clustering coefficient of the focal team with the overall or global ecosystem clustering coefficient. Therefore, we conducted a paired sample t-test to compare the local and global clustering coefficient.

Hypothesis 4 posited that scientific teams are more likely to be assembled when they are brokering ties within the local neighborhood of their team member interlock ecosystem. The paired t-test showed that the focal team's hyperedge clustering coefficient ($M = 0.61$, $SD = 0.15$) was significantly lower than the overall hypergraph clustering coefficient ($M = 0.85$, $SD = 0.01$); $t(45) = -9.86$, $p = 0.000$. The results support Hypothesis 4: Scientific teams are more likely to assemble when there is more brokerage in the local neighborhood of their team member interlock ecosystem than the global neighborhood.

Hypothesis 5 posited that scientific teams are more likely to be assembled when they are brokering ties within the local neighborhood of their team knowledge interlock ecosystem. The paired t-test showed that the focal team's hyperedge clustering coefficient ($M = 0.16$, $SD = 0.02$) is significantly lower than the hypergraph clustering coefficient ($M = 0.57$, $SD = 0.01$); $t(45) = -23.77$, $p = 0.000$. The results support Hypothesis 5: Scientific teams are more likely to assemble when there is more brokerage in the local neighborhood of their team knowledge interlock ecosystem than the global neighborhood.

Hypothesis 6 posited that scientific teams are more likely to be assembled when they are brokering ties within the local neighborhood of their team member and knowledge interlock ecosystem. The paired t-test showed that the focal team's hyperedge clustering coefficient ($M = 0.10$, $SD = 0.01$) is significantly smaller than hypergraph clustering coefficient ($M = 0.60$, $SD = 0.01$); $t(45) = -35.03$, $p = 0.000$. The results support Hypothesis 6: Scientific teams are more likely to assemble when there is more brokerage in the local neighborhood of their team member and knowledge interlock ecosystem than the global neighborhood.

We concluded our investigation of the effects of local brokerage on team assembly with the exploratory research question: Which interlock type is more important? We conducted paired sample t-tests to compare each pair of local clustering coefficient scores: team member interlocks, team knowledge interlocks, and the intersection of member and knowledge interlocks. Examining the results of these t-tests shows that the hyperedge clustering coefficient for team member interlock networks ($M = 0.61$, $SD = 0.02$) is significantly higher than the hyperedge clustering coefficient for team knowledge interlock networks ($M = 0.16$, $SD = 0.02$); $t(45) = 14.02$, $p = 0.000$ and significantly higher than the hyperedge clustering coefficient for the intersection of the team member and knowledge interlock networks ($M = 0.11$, $SD = 0.01$); $t(45) = 18.95$, $p = 0.000$. Furthermore, paired t-test showed that the hyperedge clustering coefficient for team knowledge interlock networks ($M = 0.16$, $SD = 0.02$) tends to be significantly higher than the hyperedge clustering coefficient for team member and knowledge interlock networks ($M = 0.11$, $SD = 0.01$); $t(45) = 7.122$, $p = 0.000$.

These results suggest that high local brokerage (i.e., low hyperedge clustering coefficient) in the intersection of the team member and knowledge interlock networks has a stronger effect on team assembly than local brokerage in knowledge or team member interlock networks alone.

## *Ecosystem decentralization*

Our last set of analyses examined the level of hypergraph centralization across the scientific ecosystem, and the Research Question 3: Are scientists more likely to assemble into a team in ecosystems characterized by lower levels of centralization (i.e., greater ecosystem decentralization)? Table 3 presents the hypergraph centralization for the observed ecosystems and their z-scores based on 200 simulations. A positive z-score indicates a score that is higher than expected by chance. A negative z-score indicates a score that is less than expected by chance. The larger the z-score, the greater difference there is between the score and the mean.

Hypothesis 7 posited that scientific teams are more likely to assemble within team ecosystems characterized by decentralization in team member interlocks. The simulation results showed that the observed hypergraph centralization in the member interlock network is lower than expected by chance for 44 out of the 46 team ecosystems. We further analyzed the two team ecosystems with high hypergraph centralization. Our analyses showed that the members of the focal teams were also

**Table 3.** Hypergraph centralization.

| Team Id | Team member interlock | | Team knowledge interlock | | Team member & knowledge interlock | |
|---|---|---|---|---|---|---|
| | Observed value | Z-score | Observed value | Z-score | Observed value | Z-score |
| T001 | 0.22 | −26.15*** | 0.33 | −2.72** | 0.21 | −22.35*** |
| T002 | 0.28 | −43.12*** | 0.33 | −15.17*** | 0.28 | −14.93*** |
| T003 | 0.32 | −43.14*** | 0.23 | | 0.25 | |
| T004 | 0.27 | −35.31*** | 0.30 | −2.25* | 0.20 | −5.43*** |
| T005 | 0.43 | 10.83*** | 0.33 | 4.51*** | 0.22 | 2.36* |
| T006 | 0.39 | −26.37*** | 0.11 | | 0.09 | |
| T007 | 0.36 | 23.83*** | 0.23 | −22.12*** | 0.20 | −22.26*** |
| T008 | 0.24 | −94.49*** | 0.21 | −23.11*** | 0.16 | −11.52*** |
| T009 | 0.36 | −49.94*** | 0.13 | | 0.12 | 7.44*** |
| T010 | 0.29 | −47.70*** | 0.28 | 4.28*** | 0.18 | 2.02+ |
| T011 | 0.22 | −63.85*** | 0.16 | −20.19*** | 0.12 | 45.53*** |
| T012 | 0.15 | −66.19*** | 0.22 | 12.81*** | 0.10 | |
| T013 | 0.29 | −57.87*** | 0.20 | −36.06*** | 0.07 | −23.49*** |
| T014 | 0.37 | −39.08*** | 0.16 | −25.42*** | 0.09 | 14.41*** |
| T015 | 0.31 | −73.27*** | 0.07 | −22.71*** | 0.06 | 37.68*** |
| T016 | 0.32 | −55.22*** | 0.17 | | 0.16 | |
| T017 | 0.36 | −52.26*** | 0.14 | | 0.08 | −11.57*** |
| T018 | 0.25 | −59.04*** | 0.17 | 13.32*** | 0.07 | 11.22*** |
| T019 | 0.21 | −33.74*** | 0.23 | | 0.11 | |
| T020 | 0.13 | −119.98*** | 0.12 | −37.63*** | 0.11 | 33.22*** |
| T021 | 0.37 | −82.26*** | 0.11 | 1.41 | 0.06 | |
| T022 | 0.33 | −45.04*** | 0.09 | −7.11*** | 0.03 | −9.18*** |
| T023 | 0.39 | −76.80*** | 0.15 | −16.30*** | 0.08 | 27.58*** |
| T024 | 0.22 | −74.26*** | 0.15 | −8.15*** | 0.07 | 6.90*** |
| T025 | 0.28 | −85.82*** | 0.10 | −30.85*** | 0.10 | 77.60*** |
| T026 | 0.18 | −53.61*** | 0.12 | −13.34*** | 0.03 | |
| T027 | 0.37 | −83.60*** | 0.09 | | 0.05 | 19.87*** |
| T028 | 0.20 | −78.71*** | 0.10 | −31.49*** | 0.11 | 80.59*** |
| T029 | 0.30 | −223.40*** | 0.11 | | 0.06 | −12.80*** |
| T030 | 0.18 | −63.69*** | 0.10 | −31.76*** | 0.07 | |
| T031 | 0.19 | −54.94*** | 0.08 | −28.18*** | 0.06 | 65.88*** |
| T032 | 0.30 | −85.36*** | 0.09 | −22.64*** | 0.03 | −20.48*** |
| T033 | 0.33 | −49.72*** | 0.18 | | 0.06 | −13.46*** |
| T034 | 0.14 | −65.29*** | 0.08 | −19.13*** | 0.03 | |
| T035 | 0.17 | −66.85*** | 0.09 | −27.46*** | 0.04 | 41.57*** |
| T036 | 0.21 | −89.40*** | 0.07 | −21.21*** | 0.04 | 26.21*** |
| T037 | 0.19 | −84.23*** | 0.15 | 12.70*** | 0.06 | 41.44*** |
| T038 | 0.42 | −29.00*** | 0.08 | | 0.04 | |
| T039 | 0.25 | −132.80*** | 0.12 | −18.66*** | 0.04 | −0.88 |
| T040 | 0.17 | −68.91*** | 0.11 | −18.23*** | 0.03 | 25.74*** |
| T041 | 0.13 | −92.15*** | 0.08 | −23.58*** | 0.01 | |
| T042 | 0.14 | −62.07*** | 0.15 | −6.47*** | 0.03 | −1.88+ |
| T043 | 0.17 | −62.53*** | 0.15 | −1.43 | 0.02 | −3.35** |
| T044 | 0.17 | −68.71*** | 0.14 | | 0.04 | 11.60*** |
| T045 | 0.11 | −71.89*** | 0.15 | 11.42*** | 0.02 | |
| T046 | 0.15 | −68.88*** | 0.12 | | 0.02 | |

*Notes:*
1. Teams are sorted based on hypergraph size (from small to large).
2. Z-score is based on 200 simulations. Z-score is presented only for normal distribution frequencies. + 0.10 * 0.05 ** 0.01 *** 0.001.

members of the highly central teams. So, the results support Hypothesis 7: Greater decentralization in team member interlocks across a scientific ecosystem tends to increase the likelihood of team assembly, unless the ecosystem is dominated by the members of the focal team.

Hypothesis 8 posited that scientific teams are more likely to assemble within team ecosystems characterized by greater decentralization in team knowledge interlocks. The simulation results showed that the observed hypergraph centralization in the team knowledge interlock network is lower than expected by chance for 26 out of 46 teams. For the remaining 20 teams, the frequency distribution in the null models did not follow a normal distribution, or the results showed that team knowledge interlock network is higher than expected by chance. These partial results might be explained by the fact that few teams publish interdisciplinary articles that connect to a high number of single discipline articles, thus having a high degree of centrality, which in turn generate ecosystems with high centralization.

Hypothesis 9 posited that scientific teams are more likely to assemble within team ecosystems characterized by decentralization in the intersection of the team member and knowledge interlocks. The simulation results showed that the observed hypergraph centralization in the intersection of the team member and knowledge interlock network is lower than expected by chance in only for 14 out of 46 teams. For the remaining 32 teams, the frequency distribution in the null models did not follow a normal distribution or the results showed that team member and knowledge interlock network is higher than expected by chance. Therefore Hypothesis 9 is not supported.

The results reveal the answer to our final exploratory question: *Which is most important: decentralization in a (a) team member interlock, (b) team knowledge interlock, or (c) intersection of team member and knowledge interlock ecosystem?* Given that only Hypothesis 7 was fully supported, there was no reason to proceed with paired sample t-tests comparing the effects of each pair of ecosystem centralization scores. Our results suggest that decentralization in the team member interlock network is more important to team assembly as compared to decentralization in the knowledge interlock network or the intersection of the team member and knowledge interlock network.

## Discussion

Today's most pressing problems necessitate that individuals work in teams. Social network approaches have proved valuable in providing theoretical lenses and methodological approaches to advance key questions about the assembly of teams. However, the network lens as it has been previously applied, using one and two mode networks with scientists or teams as nodes, misses important structural forces at the ecosystem level that shape the assembly of teams. Whereas some prior research has recognized this shortcoming, this article advances conceptual thinking by introducing a novel set of metrics and methods to systematically explore the multilevel forces affecting team assembly. By doing so, it makes two primary contributions.

### *Contribution #1: The effects of ecosystems on teams*

This article elucidates the significance of team interlock ecosystems, arguing that their characteristics determine the availability of unique knowledge teams use to solve complex issues, and the degree to which knowledge is shared and new ideas are generated. Therefore, understanding the drivers of team assembly requires modelling the ecosystem characteristics from which teams nucleate. There is a well-worn adage in networks research that first people make networks, but then, networks make the people (Padgett & Powell, 2012). The same can be extended to teams. First teams help us make ecosystems, but then ecosystems make the team.

We examined two components of the collaboration ecosystems that we believe are especially relevant to team assembly: team and knowledge interlocks. Our results demonstrated the potential of the larger social (member interlocks) and cognitive (knowledge network interlocks) environment in which scientists work to influence team assembly. Additionally, our depiction of scientific team interlock ecosystems incorporates the distinct effects of proximal (i.e., local) neighborhood and as well as distal (i.e., global) neighborhood effects on team assembly. We argue that this more nuanced view is necessary in order to accurately capture the various features of scientific ecosystems that play a role in team assembly.

Our findings make several substantive, albeit preliminary, contributions to the impact of ecosystems on assembly of scientific teams. First, we find that scientific teams are more likely to assemble when the *global* ecosystem formed by team member, team knowledge, and the confluence (or intersection) between membership and knowledge interlocks form a *coherent* (i.e., cohesive) intellectual ecosystem. Interestingly, we find that, as posited in Hypothesis 1, the extent to which the ecosystem exhibits teams with higher overlapping membership (as compared to overlapping knowledge or the confluence or intersection of the two) is the most important motivator for teams to assemble.

We also find that scientific teams are more likely to assemble when the *local* ecosystem exhibits higher brokerage (i.e., low hyperedge clustering coefficient) for team member and knowledge interlocks (the confluence or intersection between the two). Interestingly, the team member interlocks and team knowledge interlocks each taken by themselves were less powerful motivators for teams to assemble. We interpret these results as evidence that individuals are especially motivated to come together as a team when they feel challenged to come up with new ideas that break norms and when they perceive that their scientific endeavors will benefit from bringing diverse individuals and the accompanying resources from other teams that do not have substantial member or knowledge overlap.

Finally, when examining global ecosystem decentralization, we only found support for our hypothesis regarding team member interlocks (H7). Our findings show that the global ecosystem decentralization in team member interlocks increases the likelihood for a team to assemble. A decentralized ecosystem essentially means there is not one or a few dominant teams that have member and/or knowledge overlap with a high number of other teams. Importantly, whereas decentralization in interlocks among team members positively influenced team assembly, decentralization in knowledge interlocks, or in the confluence (or intersection) of the team member and knowledge interlock network did not motivate team assembly. Thus, the interlocks among *people* were a stronger force than the interlocks among ideas. These insights about team assembly could not be discerned using traditional network methods. And so our second primary contribution is the advancement of a relatively new approach to the study of teams (and other collectives) using network analysis based on hypergraphs.

### Contribution #2: Development of hypergraph methodology for describing and testing hypotheses about team ecosystems

We advanced the use of a *hypergraph* methodological approach, which better accounts for the nested structure of individuals in teams and the interlocks among teams with regard to knowledge and shared membership. Specifically, we model how characteristics of ecosystem coherence, local brokerage, and ecosystem decentralization affect team-based assembly. We conceptualized collaboration ecosystems as comprised of interlocking teams that overlap by virtue of shared members and shared knowledge or research topics.

Specifically, we introduced *hypergraph* measures, which better characterize the nested structure of multiple individuals and knowledge domains in multiple teams and the interlocks among them. Hypergraph approaches take sets of nodes, or hyperedges, and examine hyperties of overlapping members or knowledge domains existing between hyperedges. In our study, scientific articles (i.e., hyperedges) represent the members of a team and the knowledge areas they represent—the outcome of assembly processes in which different types of nodes (scientists and knowledge topics) are combined within an ecosystem of prior relations (i.e., hypergraph). We used the notion of local hyperedge clustering coefficient to examine the effect of brokerage within the proximal (i.e., local) neighborhood of focal team on team assembly. We used the notion of hypergraph clustering coefficient to account for the cohesion in the team's global ecosystem. And, finally, we used the notion of hypergraph centralization to describe whether the team's ecosystem is dominated by central teams.

In addition to contributing to the development of new hypergraph metrics to describe ecosystems, we also contribute new methods to test hypotheses about the impact of these metrics on team assembly. Specifically, we proposed a methodology to test hypotheses by comparing the hypergraph metrics in the observed ecosystem to those that were computed in randomly generated ecosystems that matched the observed ecosystem in terms of number of teams, number of members in teams, and number of knowledge areas in teams.

### *Limitations and future directions*

Our study advances a hypergraph approach to understanding scientific teams, and presents initial evidence documenting substantively significant effects of the local and global ecosystem on team assembly. However, there are a number of important limitations that need to be acknowledged. First, it is important to recognize that we examined the factors influencing team assembly considering only teams that had successfully assembled. We did not have access to the "invisible" collection of individuals who considered submitting a proposal but did not get around to doing so. To partially address this limitation, we compared the observed ecosystem of the assembled team with a set of randomly simulated ecosystems. Future research should explore the development of an analytic approach that compares the ecosystem of an assembled team with the observed ecosystem of a random group of researchers, matched on some key characteristics, who never assembled into a team.

Second, the ecosystem of teams used in this study was created based on a set of focal teams who submitted research proposals to a specific grant competition. There are likely to be specifics of the domain and competition that affect the nature of the ecosystem. Thus, it is important for future work to continue to explore other types of collaborative ecosystem contexts. For example, it would be valuable for future research to explore the effects of policy interventions like the creation of centers or calls for learning communities or research coordination networks as discontinuities in how teams assemble. Furthermore, future research should examine whether the results are supported in other contexts. Our study examined the influence of team ecosystem structures on team assembly in the area of clinical and translational science. It would be important to identify the ecosystem structures influencing the assembly of teams and their subsequent team interactions within other areas of research—and indeed beyond scientific collaboration to other contexts where teams are increasingly being self-assembled to engage in critical tasks.

A third limitation concerns the creation of our knowledge network. We built our database of keywords by identifying those keywords contained in the proposal of at least one team. However, the entirety of available knowledge circulating within the ecosystems surrounding a focal team would likely contain keywords that were in articles not written by these proposal teams. Future work on knowledge interlocks should consider alternative methods of building knowledge networks.

A fourth limitation of this study is that we did not examine the *consequences* of team assembly. The scope of this article was to develop the methodology to examine ecosystem forces driving the assembly of teams. An interesting next question is, which team assembly mechanisms are beneficial or detrimental in terms of the ultimate creativity and innovation produced by a new team. At the ecosystem level, comparing the characteristics of multiple ecosystems could allow researchers to explore questions of which ecosystem characteristics are more or less functional to spawning the assembly of innovative teams. Although these cross-ecosystem comparisons were beyond the scope of this article, the methodology developed here can easily be adapted to answer questions such as: Which ecosystem factors are likely to explain why most teams form? And, are there different ecosystem factors that explain why only some of those teams perform effectively?

A final limitation is the context of studying assembly in response to a call for funding proposals. First, in some areas, funding does not play a major role in scientific advancement and the insights from this study may not generalize to those areas. In the area examined here—clinical and translational science—there is a very strong reliance on funding. But even in this case, the call for proposals was for a relatively small amount of seed funding. It is possible that the ecosystem included people who were already well funded and in established teams who would not be motivated to submit a proposal even if the domain of their research is well aligned with the call for proposals. It is noteworthy that we found significant ecosystem effects on team assembly despite the aforementioned reasons why some might not be motivated to assemble into teams to submit a proposal.

## Conclusion

Given the increasing importance of teams for innovation, research is needed to uncover the factors that shape the assembly of teams in domains like scientific knowledge production. This article advances this area by introducing new hypergraph metrics and simulation -methodologies for inferentially testing hypotheses about the impact of team and knowledge interlocks on team assembly. These efforts will pave the way for conceptual advances that have been called for by practitioners as well as researchers studying teams, but heretofore, have remained largely unexplored theoretically and empirically.

## Funding

The preparation of this manuscript was supported by funding from the Army Research Office [ARO W911NF-14-10686] and the National Institutes of Health [R01GM112938, U01GM112623, UL1RR025741].

## ORCID

Alina Lungeanu http://orcid.org/0000-0003-3368-6339

## References

Acedo, F. J., Barroso, C., Casanueva, C., & Galán, J. L. (2006). Co authorship in management and organizational studies: An empirical and network analysis. *Journal of Management Studies*, *43*(5), 957–983. doi:10.1111/j.1467-6486.2006.00625.x

Barabási, A.-L. (2005). Network theory–The emergence of the creative enterprise. *Science*, *308*(5722), 639–641. doi:10.1126/science.1112554

Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*(5439), 509–512. doi:10.1126/science.286.5439.509

Berge, C. (1973). Graphs and hypergraphs. Amsterdam, The Netherlands: North-Holland.

Bordons, M., & Gómez, I. (2000). Collaboration networks in science. *The web of knowledge: A festschrift in honor of Eugene Garfield*. Medford, NJ: Information Today, 197–213.

Bulmer, M. G. (1979). *Principles of statistics*. New York, NY: Dover Publications, Inc.

Burt, R. S. (1992). *Structural holes: The social structure of competition*. Cambridge, MA: Harvard University Press.

Charney, D. (2003). Lone geniuses in popular science: The devaluation of scientific consensus. *Written Communication*, *20*(3), 215–241. doi:10.1177/0741088303257505

Cummings, J. N., & Kiesler, S. (2007). Coordination costs and project outcomes in multi-university collaborations. *Research Policy*, *36*(10), 1620–1634. doi:10.1016/j.respol.2007.09.001

Cummings, J. N., & Kiesler, S. (2008). *Who collaborates successfully?: Prior experience reduces collaboration barriers in distributed interdisciplinary research*. Paper presented at the Proceedings of the 2008 ACM conference on Computer supported cooperative work, San Diego, CA, USA.

De Solla Price, D. J. (1965). Networks of scientific papers. *Science*, *149*(3683), 510–515. doi:10.1126/science.149.3683.510

Dunbar, R. I. (1992). Neocortex size as a constraint on group size in primates. *Journal of Human Evolution*, *22*(6), 469–493. doi:10.1016/0047-2484(92)90081-J

Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, *44*(2), 350–383. doi:10.2307/2666999

Estrada, E., & Rodríguez-Velázquez, J. A. (2006). Subgraph centrality and clustering in complex hyper-networks. *Physica A: Statistical Mechanics and its Applications*, *364*, 581–594. doi:10.1016/j.physa.2005.12.002

Fleming, L., Mingo, S., & Chen, D. (2007). Collaborative brokerage, generative creativity, and creative success. *Administrative Science Quarterly*, *52*(3), 443–475. doi:10.2189/asqu.52.3.443

Ghasemian, F., Zamanifar, K., & Ghasem-Aghaee, N. (2017). Composing scientific collaborations based on scholars' rank in hypergraph. *Information Systems Frontiers*, 1–16.

Ghasemian, F., Zamanifar, K., Ghasem-Aqaee, N., & Contractor, N. (2016). Toward a better scientific collaboration success prediction model through the feature space expansion. *Scientometrics*, *108*(2), 777–801. doi:10.1007/s11192-016-1999-x

Ghoshal, G., Zlatic, V., Caldarelli, G., & Newman, M. E. J. (2009). Random hypergraphs and their applications. *Physical Review E*, *79*(6), 066118. doi:10.1103/PhysRevE.79.066118
Grant, R. M. (1996). Toward a knowledge-based theory of the firm. *Strategic Management Journal*, *17*, 109–122. doi:10.1002/smj.4250171110
Gruenfeld, D. H., Mannix, E. A., Williams, K. Y., & Neale, M. A. (1996). Group composition and decision making: How member familiarity and information distribution affect process and performance. *Organizational Behavior and Human Decision Processes*, *67*, 1–15. doi:10.1006/obhd.1996.0061
Guimera, R., Uzzi, B., Spiro, J., & Amaral, L. A. N. (2005). Team assembly mechanisms determine collaboration network structure and team performance. *Science*, *308*(5722), 697–702. doi:10.1126/science.1106340
Heider, F. (1958). *The psychology of interpersonal relations*. Hillsdale, NJ: Lawrence Erlbaum Associates.
Hinds, P. J., Carley, K. M., Krackhardt, D., & Wholey, D. (2000). Choosing work group members: Balancing similarity, competence, and familiarity. *Organizational Behavior and Human Decision Processes*, *81*(2), 226–251. doi:10.1006/obhd.1999.2875
Kozlowski, S. W., & Klein, K. J. (2000). *A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes*. In K. J. Klein & S. W. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions*. San Francisco, CA: Jossey-Bass.
Kuhn, T. S. (1996). *The structure of scientific revolutions* (3rd ed.). Chicago, IL: University of Chicago Press.
Lee, S., & Bozeman, B. (2005). The impact of research collaboration on scientific productivity. *Social Studies of Science*, *35*(5), 673–702. doi:10.1177/0306312705052359
Lungeanu, A., & Contractor, N. S. (2015). The effects of diversity and network ties on innovations: The emergence of a new scientific field. *American Behavioral Scientist*, *59*(5), 548–564. doi:10.1177/0002764214556804
Lungeanu, A., Huang, Y., & Contractor, N. S. (2014). Understanding the assembly of interdisciplinary teams and its impact on performance. *Journal of Informetrics*, *8*(1), 59–70. doi:10.1016/j.joi.2013.10.006
Lungeanu, A., Sullivan, S., Wilensky, U., & Contractor, N. S. (2015). *A computational model of team assembly in emerging scientific fields*. Paper presented at the Winter Simulation Conference, Huntington Beach, California.
McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, *27*, 415–444. doi:10.1146/annurev.soc.27.1.415
Moody, J. (2004). The structure of a social science collaboration network: Disciplinary cohesion from 1963 to 1999. *American Sociological Review*, *69*(2), 213–238. doi:10.1177/000312240406900204
Murray, F., & O'Mahony, S. (2007). Exploring the foundations of cumulative innovation: Implications for organization science. *Organization Science*, *18*(6), 1006–1021. doi:10.1287/orsc.1070.0325
National Research Council. (2015). *Enhancing the effectiveness of team science*. Washington, DC: National Academies Press.
Newman, M. E. J. (2001). Clustering and preferential attachment in growing networks. *Physical Review E*, *64*(2), 025102. doi:10.1103/PhysRevE.64.025102
O'Leary, M. B., Mortensen, M., & Woolley, A. W. (2011). Multiple team membership: A theoretical model of its effects on productivity and learning for individuals and teams. *Academy of Management Review*, *36*(3), 461–478.
Oh, H., Chung, M.-H., & Labianca, G. (2004). Group social capital and group effectiveness: The role of informal socializing ties. *Academy of Management Journal*, *47*(6), 860–875. doi:10.2307/20159627
Oh, H., Labianca, G., & Chung, M.-H. (2006). A multilevel model of group social capital. *Academy of Management Review*, *31*(3), 569–582. doi:10.5465/AMR.2006.21318918
Padgett, J. F., & Powell, W. W. (2012). *The emergence of organizations and markets*. Princeton, NJ: Princeton University Press.
Poole, M. S., & Contractor, N. S. (2011). Conceptualizing the multiteam system as an ecosystem of networked groups. In S. J. Zaccaro, M. A. Marks, & L. A. DeChurch (Eds.), *Multiteam systems: An organizational form for dynamic and complex environments*. New York: Routledge Academic.
Reagans, R., Zuckerman, E., & McEvily, B. (2004). How to make the team: Social networks vs. demography as criteria for designing effective teams. *Administrative Science Quarterly*, *49*(1), 101–133.
Sharma, A., Srivastava, J., & Chandra, A. (2014). *Predicting multi-actor collaborations using hypergraphs*. ArXiv eprints, 1401.6404.
Shi, F., Foster, J. G., & Evans, J. A. (2015). Weaving the fabric of science: Dynamic network models of science's unfolding structure. *Social Networks*, *43*, 73–85. doi:10.1016/j.socnet.2015.02.006
Sullivan, S., Zhu, M., Lungeanu, A., & Contractor, N. S. (2012). *Hypergraph metrics*. Evanston, IL: Northwestern University.
Taramasco, C., Cointet, J. P., & Roth, C. (2010). Academic team formation as evolving hypergraphs. *Scientometrics*, *85*, 721–740. doi:10.1007/s11192-010-0226-4
Torvik, V. I., Weeber, M., Swanson, D. R., & Smalheiser, N. R. (2005). A probabilistic similarity metric for Medline records: A model for author name disambiguation. *Journal of the Association for Information Science and Technology*, *56*(2), 140–158. doi:10.1002/asi.20105
Uzzi, B., & Spiro, J. (2005). Collaboration and creativity: The small world problem. *American Journal of Sociology*, *111*(2), 447–504. doi:10.1086/432782

Wang, J.-W., Rong, -L.-L., Deng, Q.-H., & Zhang, J.-Y. (2010). Evolving hypernetwork model. *The European Physical Journal B - Condensed Matter and Complex Systems*, *77*(4), 493–498. doi:10.1140/epjb/e2010-00297-8

West, M. A., & Anderson, N. R. (1996). Innovation in top management teams. *Journal of Applied Psychology*, *81*(6), 680–693. doi:10.1037/0021-9010.81.6.680

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, *316*(5827), 1036–1039. doi:10.1126/science.1136099

Zaheer, A., & Soda, G. (2009). Network evolution: The origins of structural holes. *Administrative Science Quarterly*, *54* (1), 1–31. doi:10.2189/asqu.2009.54.1.1

Zlatic, V., Ghoshal, G., & Caldarelli, G. (2009). Hypergraph topological quantities for tagged social networks. *Physical Review E*, *80*(3), 036118. doi:10.1103/PhysRevE.80.036118