ISBN

978-3-319-72149-1

Authors

Jones, Alex T
Friedkin, Noah E
Singh, Ambuj K

Publication Date

2018

DOI

10.1007/978-3-319-72150-7_71

Peer reviewed

Modeling the Co-evolution of Committee Formation and Awareness Networks in Organizations

Alex T. Jones[1(✉)], Noah E. Friedkin[2], and Ambuj K. Singh[1]

[1] Department of Computer Science, University of California, Santa Barbara, CA, USA
{alexjones,ambuj}@ucsb.edu
[2] Department of Sociology and the Center for Control, Dynamical-Systems and Computation, University of California, Santa Barbara, CA, USA
friedkin@soc.ucsb.edu

Abstract. Large-scale organizations confront a difficult problem of assigning personnel to non-routine tasks. Such projects require novel combinations of individuals from separate parts of the organization. We propose a computational method for assembling a committee that recruits individuals to solve the task. This method is driven by data on interpersonal awareness relationships, fostered by mutual interaction and indirect contact. By inducing a network from these relationships, we can reduce the committee selection problem to the maximum coverage problem. A stochastic model is then presented to capture plausible new relationships which form during committee deliberation. This framework provides administrators the opportunity to track the evolution of awareness relationships while improving the organization's ability to solve non-routine tasks. To track awareness network evolution, we demonstrate correlations in empirical networks between communication network topology and the probability of awareness relationships. To formalize organizational improvement over a series of committee formations, we extend a static committee formation problem to a repeated committee formation problem. After showing the computational cost of exact solutions to this problem, we propose an efficient, heuristic-based solution. Through simulation on real-world networks, we demonstrate that our approach produces more efficient and productive network states than a baseline algorithm during a series of committee formations.

1 Introduction

In large-scale organizations, assignments to non-routine projects are often handled by small groups of decision makers. Our focus is on the extent to which they, collectively, possess sufficient information to intelligently decide which members of the organization should be assigned to a project. A committee's potential to select a successful team depends on the scope and accuracy of the assigners' pool

C. Cherifi et al. (eds.), *Complex Networks & Their Applications VI*,
Studies in Computational Intelligence 689, https://doi.org/10.1007/978-3-319-72150-7_71

of information on the performances, circumstances, and expected behaviors of other individuals.

We approach this problem through study of two key networks present in formal, task-oriented organizations. First, individuals establish communication channels with other individuals and form contact relationships. Over these channels, individuals accumulate and store information about individuals' past and current activities, successes and failures, abilities and weaknesses. These relationships induce a *contact network*. Interaction through communication channels updates the memory records of the interacting individuals about one another and other individuals. The pool of all such memory records on a particular individual contains far more information about the individual than is available from any archive of written evaluations. We construct an *awareness network* induced by edges (i, j) when i possess a memory record, that is information on current work-related activities, of j. Here, we investigate relationships between these two networks and the resulting impact on non-routine task assignment.

Our first contribution is reducing the committee selection problem to the maximum coverage problem in the awareness network, a classic graph-theoretic problem. This is accomplished by demonstrating that an optimal non-routine task assignment committee is a group of individuals with the largest pool of unique memory records. Using an estimate of the awareness network (via survey or other methods), this permits a logical, data-driven method for administrators to select committees.

Our second contribution is modeling new contact relationships formed due to group discussion in the committee. Committee deliberation leads to contact network *augmentation*, the addition of new edges to the network. This creates new information channels that lead to acquisition of personal information records by committee members. We describe this augmentation process using a simple stochastic model.

To test the flexibility of our tools we answer the following question: can a committee be chosen to assign a team to the task at hand *and* improve the efficiency and productivity of the organization? Our third contribution addresses this question by studying the extension of the static committee formation problem to the repeated committee formation problem. This problem involves selecting and forming a sequence of committees, intending to maximize the quality of the terminal committee. We discuss the problem of computing optimal committee sequences and provide an efficient, heuristic-based algorithm called the Diverse Coverage (DC) algorithm. We show that our methodology avoids pitfalls of naive, greedy methods.

Our forth and final contribution is a longitudinal simulation comparing the DC algorithm to a baseline algorithm. Our networks are initialized using survey data gathered in [6]. We first demonstrate that repeated committee formation improves coverage quality over time. Our results show that the DC algorithm simultaneously provides high quality committees at each step, increases efficiency and productivity in the organization, and increases the committee quality over time.

We organize our paper as follows. First, we discuss contact and awareness networks in organizations and their relationship to one another. Next, we demonstrate how to select an optimal committee using these networks. We then motivate and describe a multi-step optimization problem designed to improve committee quality over time. After developing a solution to this problem, we demonstrate its effectiveness through simulations.

Related Work

Research on transactive memory systems (TMS) has shown that the communication networks of task-oriented small groups allow their members to informally acquire information about one another and settle upon their division of labor [12,16,21]. Here, we extend the scope of TMS to cover a large group whose members are connected by a contact network. Further, researchers have noted a lack in understanding about how TMS evolve in dynamic settings [19]. Our results further the understanding of TMS evolution when repeatedly executing non-routine tasks.

Researchers have developed theoretical tools in order to directly solve task assignment and team selection problems. The work of [3] applies coordination theory to redesign assignment processes in a software bug-fixing pipeline. Alternatively, a graph-theoretic approach has shown that knowing what tasks each individual can complete can be used to compute a satisfying assignment of workers to jobs [10]. Here, we utilize interpersonal relationship data to reason about which group of individuals can best assign a team to solve a task. This leverages existing TMS within an organization to solve non-routine tasks, rather than assigning a team directly.

Research in social networks has explored triadic closure models to explain and predict network evolution over time [8,13]. Detailed studies based on maximum-likelihood models have led to generative models of network evolution in online social networks [11]. Other social network research proposes a general actor-based stochastic model for capturing network evolution dynamics [20]. Here, we draw on triadic closure dynamics to design an augmentation regimen that improves organization performance.

Literature on cognitive social networks has explored relationships between social structure and underlying cognitive relationships [2,9]. Here, the cognitive networks are induced by work-related information records, rather than on perceived social relationships

Research in coupled social networks provides evidence that behavioral relationships precede cognitive relationships [1]. This strengthens our assumption that contact relationships, rather than awareness relationships, form directly due to committee deliberation. Social network research has also shown that action around foci increases the likelihood of social ties [5]. Further, the confluence of social proximity, geographic proximity, and shared social context encourages personal relationships [15]. These results are consistent with our augmentation model, as committee discussion events fit the necessary criteria for new relationships to form.

2 Contact and Awareness Networks

We use the terms **augmentation** of a graph as the addition of edges to the graph and **augmented graph** as the new graph produced by this process. The **coverage** of a node set S in a graph G is the number of nodes which are reachable in at most one-hop from S, denoted $\bar{\mathcal{N}}_G(S)$, and a node i is **covered** by set S if $i \in \bar{\mathcal{N}}_G(S)$.

We begin by considering the relationship between a *behavioral* network, generated by interpersonal activities, and a *cognitive* network, generated by interpersonal information on others, within an organization. The corresponding networks relevant to non-routine task assignment are the contact network and the awareness network, resp., which we describe below.

An organization's **contact network** captures all person-to-person communication channels. An edge (i, j) exists in the contact network if and only if individual i and j engage in regular discussion about work-related matters. These relationships are important because direct contact on work-related matters automatically generates memory records about personal information. Without direct contact, a memory record may still exist if information about person j has been conveyed to person i by other individuals. This implies that contact network topology influences the location of memory records in the organization.

Memory record data is important because it allows us to judge if a group of individuals can determine another individual's ability to solve a non-routine task. We define an (i, j) **memory record** to be the presence of up-to-date knowledge at individual i regarding work-related activity of j. The set of all memory records induces an **awareness network**.

The coupling between awareness and contact networks can be understood through examination of empirical evidence. The probability of a memory record is conditioned by the local contact network topology connecting the (i, j) pair. Notably, a "horizon of observability" exists. The probability that individuals i and j know anything about the work of each other is near zero when $dis(i, j) > 2$ in the contact network, where $dis(i, j)$ is the length of the shortest path between i and j. This horizon was discovered in a study of the work-related contact networks among faculty of The University of Chicago and Columbia University [6]. Within the $dis(i, j) \leq 2$ "horizon of observability", the probability of an (i, j) memory record is a function of the number of mutual contacts, as shown in Fig. 1.

To learn an explicit coupling between an organization's contact network and its awareness network, we fit a function of the likelihood of an awareness edge given the number of mutual contacts for $dis(i, j) \in \{1, 2\}$ pairs, shown in Fig. 1. Borrowing from [6], we fit curves of the form $\alpha - \beta(1 - r)^m$ to our data where m is the number of mutual contacts and r is a learned parameter. The constants α and β are learned separately for distance 1 and 2 node pairs. The functions $p^1_{ij}(m) = 1 - 0.58(1 - 0.05)^m$ and $p^2_{ij}(m) = 0.3 - 0.3(1 - 0.05)^m$ fit our data well for distance 1 and 2 pairs (i, j), resp.. Once we have all p_{ij} values, we can impute an awareness network where edge (i, j) exists with probability p_{ij}.

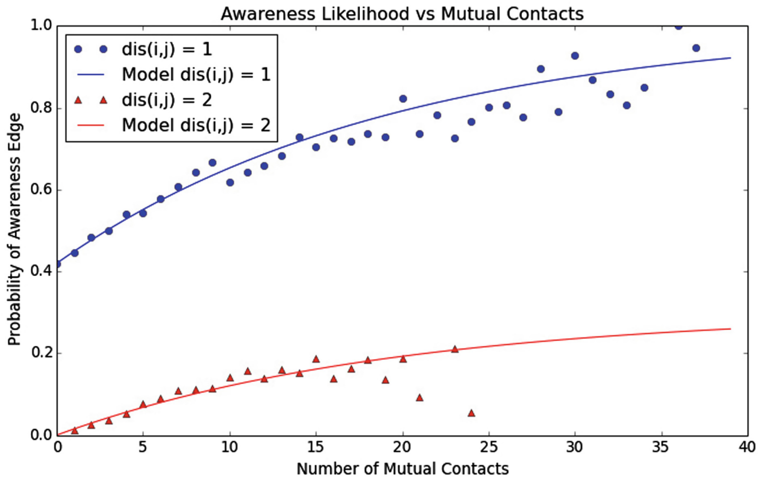


Fig. 1. The empirical likelihood of a memory record given the number of mutual contacts is drawn in blue circles and red triangles for distance one and two pairs, resp. Model fitting curves are drawn in blue and green, resp. Data are aggregated from six networks within University of Chicago and Columbia University [6]. Each proportion based on at least thirty cases.

3 Computing Committees for Non-routine Task Assignment

In this section, we address what qualities determine an optimal committee and how to compute such a committee. We then analyze how committee deliberation augments the structure of the organization's contact and awareness networks. Iterating this process, that is repeatedly forming committees, will induce dynamics on the contact network. We discuss how a forward-thinking administrator can leverage this to drive global network changes that improve organizational efficiency and productivity. We frame long-term optimization goals by describing a repeated committee formation problem. We then demonstrate the intractability of finding an exact solution and propose an efficient, heuristic-based algorithm.

3.1 Optimal Committee Selection

Suppose that a decision is required on which faculty in a university should be invited to join a new interdisciplinary institute with the mission of advancing research on a particular class of problems. Who should be recruited, and what committee is best suited to make recruitment decisions that exploit available information on the skills, current interests, behaviors, and circumstances of the faculty? Since a non-routine task may require the skills of any individual with similar probability, the committee that maintains memory records about the largest number of individuals has the greatest chance of remembering appropriate individuals to assign to the task-solving team. The discussions of such a committee will consequently be the most informed and effective at assessing the strengths and weaknesses of potential team members. Given that committees should be small, say k individuals, this committee is a solution to the *maximum coverage problem* in the awareness graph. Although calculating an exact solution to this problem is NP-Hard [4], a simple, efficient $(1 - \frac{1}{e})$ approximation exists following the classic greedy selection strategy of [17].

3.2 Committee Deliberation and Network Augmentation

A committee engages k individuals to discuss and regularly meet for a length of time. Committee members will form contact relationships with other members of the committee. This establishes the first augmentation step: add edges so that committee members form a clique in the contact network.

Naturally, active engagement between committee members may spawn new contacts; however, preexisting social and organization structures will guide the formation of indirect contacts. It is well known that triads form one of the simplest subunits of group social dynamics [5,7,8,13,15]. Open triads which include committee members may close due to committee deliberation. Consider an illustrative scenario: Professor X and Professor Y, previously not in contact, join a committee. Professor X discovers that her contact Professor Z has a project whose results directly benefit the research of Y. Consequently, X sets up a lunch meeting with Y and Z, establishing a mutual contact relationship. This relationship between Y and Z formed indirectly as a result of face-to-face interaction between committee members X and Y. We assume these edges are formed stochastically and with an equal probability. This establishes the second aug-

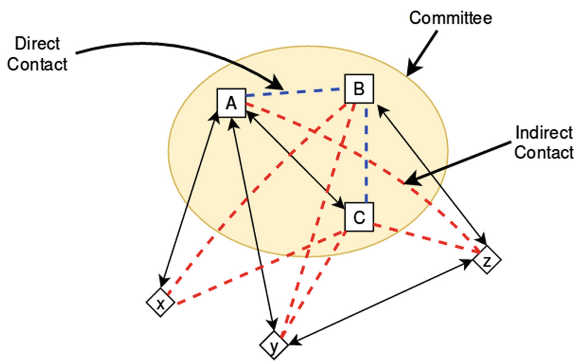


Fig. 2. Existing contact network edges are bidirectional and black. After the three person committee $\{A, B, C\}$ is formed, type I edges (blue dashed lines) $(A, B), (B, C)$ are formed between all committee pairs. Type II edges (red dashed lines) are formed with probability q due to triadic closure.

mentation step: each unclosed triad with two committee members is closed with probability q.

Summarizing the above discussion, the following steps occur during the deliberation of a committee C within the organization's existing contact network G. For simplicity, we assume edges of our contact network are bidirectional.

- **Type I Edges:** The edges $C \times C$ are added to $E(G)$.
- **Type II Edges:** For each unclosed triad x, y, z with $x, y \in C$ and $z \in \mathcal{N}_G(C)$, the absent edge is added to $E(G)$ with probability q.

In Fig. 2 we draw type I (blue) and type II (red) edges that can form when a committee $\{A, B, C\}$ is chosen.

3.3 The Repeated Committee Formation Problem

A natural goal in repeatedly forming committees is to maximize the coverage quality of the t-th committee after $t-1$ successive committee formations. We call this the **repeated committee formation problem**. An administrator which chooses committees that solve this problem maximizes the capability of the organization to respond to future non-routine tasks.

Although a maximum coverage set in the awareness network provides the highest quality committee, the resulting augmentation may not maximally improve the coverage of *future* committees. To demonstrate this, we construct an example network in the left half of Fig. 3. We iteratively select a 2-person committee by greedily choosing the optimal one-step committee, the set with highest awareness coverage. The limiting result of this process is shown in the right half of Fig. 3. Notice that the red nodes never gain any new edges. Why? Augmentation improves the coverage of the initial committee, the green nodes $\{A, B\}$, more than other committees, so $\{A, B\}$ is chosen at all remaining time steps. This demonstrates that greedily choosing the best committee may prevent the ability of future committees from covering all of network.

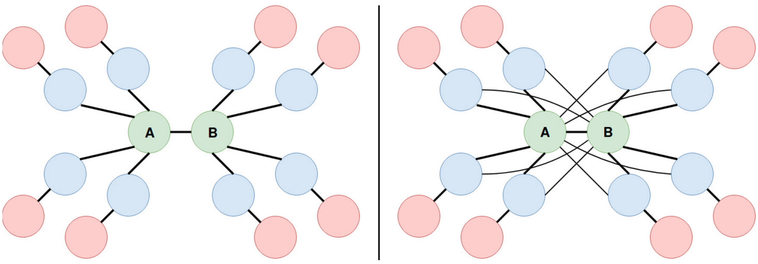


Fig. 3. The left network is an initial example contact network. The right contact network is the limiting result of greedily forming an optimal one-step committee.

3.4 Proposed Solution

It is possible to use existing algorithms to solve the repeated committee formation problem. One such approach is to construct an equivalent Markov Decision Process and use dynamic programming to solve the associated Bellman Equation [18]. Unfortunately, this method is intractable because the number of possible solutions is $O(n^{kt})$, where n is the number of members of the organization, k is the committee size, and t is the number of possible committees. In practice, we may not need to calculate an exact optimal solution. We propose the Diverse Coverage Algorithm (DC), which selects high coverage committees that will increase coverage of future committees. Our main idea is to choose a committee which, under augmentation, expands the contact neighborhood of a subset of the optimal one-step committee. Adding these contact edges immediately increases awareness and enables future committees to form more type II edges under augmentation. This strategy also bypasses the local optima shown in Fig. 3. Finally, it allows us to reason about the long-term implications of committee formation without the expense of explicitly *looking ahead* via enumeration or Monte Carlo simulation. We give psuedocode for DC in Algorithm 1 and give details in the following paragraphs.

Let $f(G) = G'$ map a contact network G to an awareness network G' whose edges exist with probability p_{ij} given by the function learned on empirical network data (see the end of Sect. 2). In this imputed network, an optimal committee is a set of nodes with maximum coverage in expectation. Solving this problem, the *expected maximum coverage problem*, is NP-Hard, but a similar efficient $(1 - \frac{1}{e})$ approximation exists (we omit the proof due to space constraints).

Notation: $G \setminus S$ represents the induced subgraph of G without the nodes in S. The inclusive neighborhood of a set of nodes B in G is denoted $\bar{\mathcal{N}}_G(B)$. The abbreviations $MC(G, k)$ and $EMC(G, k)$ refer to solutions of the Maximum Coverage Problem and Expected Maximum Coverage Problem on G with set size k, resp.

The DC algorithm proceeds in two phases and utilizes information from both the contact network and the awareness network. In the first phase, we find a seed set of size $s \leq k$, called B, with maximum expected coverage in the *awareness* network G'. This set forms the core of the committee. In the second phase, we find a diversifying set C of size $k - s$ with maximum coverage in the *contact* subnetwork H, formed by removing the inclusive neighborhood of B from the contact network. Then, we check that the utility of C exceeds a minimum threshold. To measure the utility of C, we set w to be the minimum reduction of coverage of C in H when removing any one node of C. If w is less than a parameter μ, we increment the size of the seed set and repeat the selection process. The algorithm finally returns the union of the two disjoint sets B and C as the chosen committee.

The motivation for our heuristic is as follows. The number of unique individuals that can be reached with a contact edge from B is fixed if B is the entire committee. Hence, including C in the committee to cover diverse contacts, that is nodes in H, expands the contact neighborhood of B through type II edges.

Algorithm 1: Diverse Coverage algorithm for committee selection (DC)

Data: Contact Network $G = (V, E)$, Initial seed size s, Committee size k, Minimum utility μ
Result: $S \subseteq V$
$w \leftarrow \infty$
$G' = f(G)$
while $w \geq \mu$ *and* $s \leq k$ **do**
  $B = EMC(G', k)$
  $H = G \setminus \bar{\mathcal{N}}_G(B)$
  $C = MC(H, k - s)$
  $w = \min_{v \in C} \left[|\bar{\mathcal{N}}_H(C)| - |\bar{\mathcal{N}}_H(C \setminus v)| \right]$
  $s = s + 1$
end
$S = B \cup C$

As a result, the utility of the diversifying set C will eventually fall below the threshold and s will increase, improving current committee quality. The value of μ modulates the rate at which our committee quality improves, with lower values providing slower improvement but ultimately higher coverage values.

4 Simulation of Repeated Committee Formation

In this section, we analyze network dynamics when simulating repeated committee formation using (i) the DC algorithm and (ii) randomly choosing a committee. The random method represents an administrator choosing individuals based on attributes independent of awareness network topology, which we use as a baseline. We initialize four different contact networks from the survey data gathered from networks in the Biological Sciences and Physical Sciences Divisions at University of Chicago and Columbia University. One iteration of the simulation is as follows. Using the functions in functions given in Sect. 2, we estimate an awareness network from the current contact network. We select a committee on this estimated network and augment the contact network according to our model. The augmented contact network is used as the contact network in the next iteration.

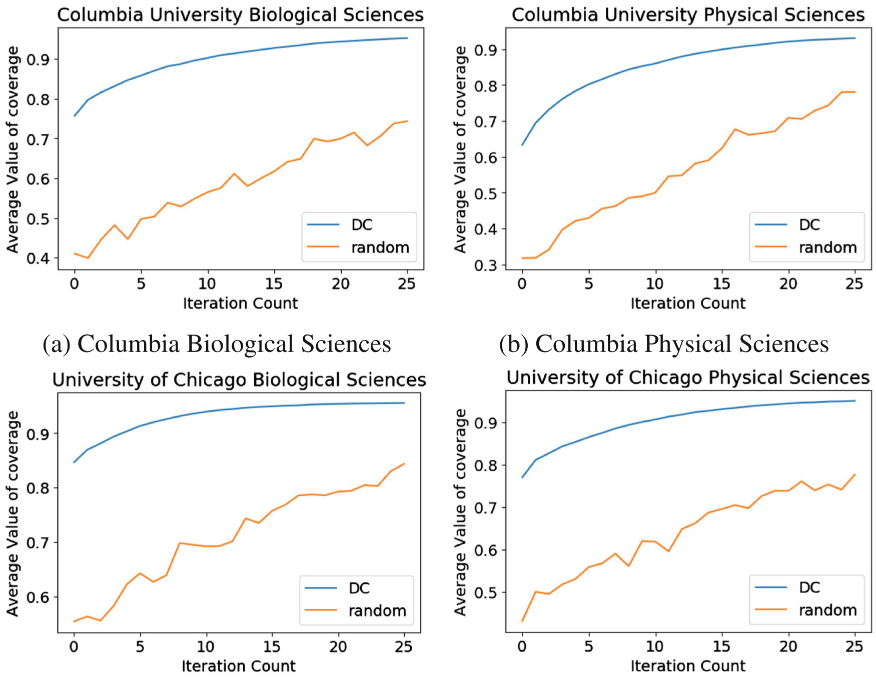
Our simulation results are averaged over multiple trials, each trial simulating 25 committee formations. Our networks contain between 105–157 vertices, so $k = 6$ is chosen to avoid too low or too high initial committee quality. We found choosing $s = 4$ and $\mu = 1$ balances quick improvement in coverage while providing high terminal coverage. Our results hold over various values of triadic closure probability, denoted q, but choosing $q = 0.08$ results in roughly 10 new contacts per person forming during the first augmentation step, which we deem plausible. These parameter values are chosen for all simulations that follow.

4.1 Effects on Committee Quality

Compared with randomly chosen committees, DC committees have markedly higher expected coverage values. This gap is between 22 and 30% of the network initially. The gap in coverage quality between these two policies remains large even after many iterations. In Fig. 4a–d, we track the expected number of covered nodes in the awareness network for each committee. Also, we calculated an approximately optimal committee for each graph created during the augmentation processes. The DC algorithm increases the maximum expected coverage value faster than the random solution, with a maximum gap between 3–10% for our networks.

4.2 Effect on Information Flow

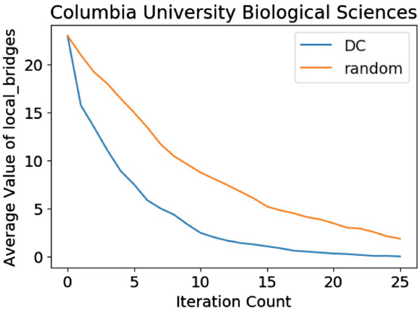
Repeated committee formation increases the capacity of information channels in the network. To measure bottlenecks of information flow, we track the number of local bridges present in the contact network. A local bridge exists when edge (i, j) exists but no length 2 path connects i and j. These edges represent bottlenecks of



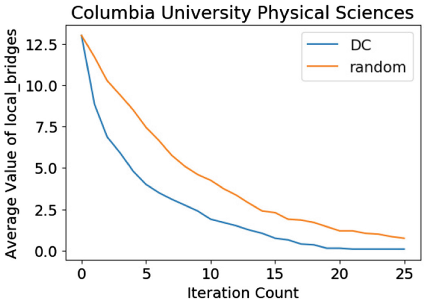
(a) Columbia Biological Sciences (b) Columbia Physical Sciences

(c) Chicago Biological Sciences (d) Chicago Physical Sciences

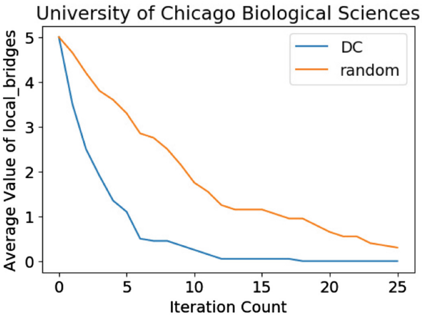
Fig. 4. Expected committee coverage in estimated awareness networks. The DC method selects committees with significantly higher expected coverage at all times.



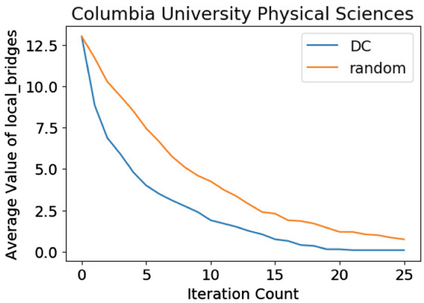
(a) Columbia Biological Sciences



(b) Columbia Physical Sciences



(c) Chicago Biological Sciences

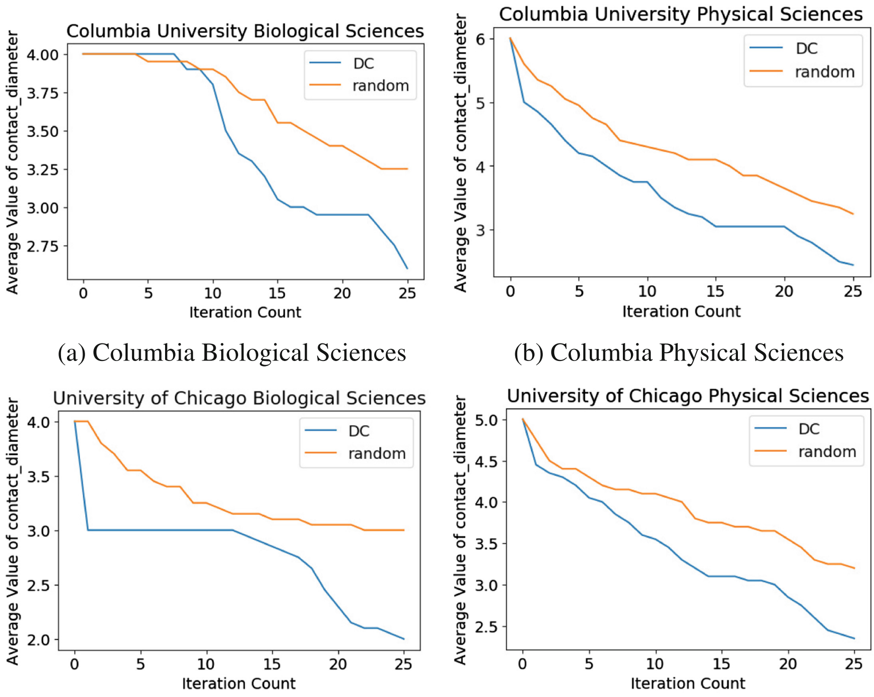


(d) Chicago Physical Sciences

Fig. 5. Local bridge count in contact networks. The DC method results in a smaller number of bridges.

information in the network, as they often form the only plausible communication path between neighbors of i and neighbors of j [7]. Our method reduces the number of local bridges in the network, leading to freer exchange of information between the initially differentiated components of the network. The reduction under the DC method outperforms the random method at all times, see Fig. 5a–d. This result can be expected because the DC method merges communities into a single structure, bypassing existing information bottlenecks at the local bridges.

Another important measure of information exchange efficiency is the network diameter. A smaller contact network diameter implies a direct improvement in referral efficiency and reduces the delay of word-of-mouth information propagation. The DC method reduces the diameter faster than the random method. See Fig. 6a–d. In the limit, the DC strategy will bring at least $|V| - \mu$ nodes into the inclusive contact neighborhood of the committee and produce a contact network diameter of at most $1 + \mu$. It is important to note that as q is increased, the expected number of iterations required to reach this limit is reduced.



(a) Columbia Biological Sciences

(b) Columbia Physical Sciences

(c) Chicago Biological Sciences

(d) Chicago Physical Sciences

Fig. 6. Contact network diameter decreases faster under the DC method.

5 Discussion

Our analysis and model highlight a new computational approach for decision making in organizations. Our method is grounded by establishing a relationship between contact and awareness network topology in organizations. This enables us to compute an optimal committee for non-routine task assignments. We then demonstrate that committee deliberation can improve committee quality over time by augmenting contact and awareness networks. Since optimizing these choices is intractable, we propose an efficient, heuristic-based algorithm for repeatedly selecting committees. Simulation on real-world data demonstrates our algorithm creates more efficient and productive network states than a baseline approach that lacks knowledge of the awareness network. By operating with an understanding of an organization's TMS, we demonstrate short-term committee quality goals and beneficial long-term structural changes can be simultaneously achieved.

Acknowledgement. This work was supported by the U. S. Army Research Laboratory and the U.S. Army Research Office under grant number W911NF-15-1-0577 and the National Science Foundation under grant number DGE-1258507.

References

1. Bahulkar, A., et al.: Co-Evolution of two networks representing different social relations in netsense. Stud. Comput. Intell. Complex Netw. Appl. V **2016**, 423–434 (2016)
2. Carley, K.: An approach for relating social structure to cognitive structure. J. Math. Sociol. **12**(2), 137–189 (1986)
3. Crowston, K.: A coordination theory approach to organizational process design. Organ. Sci. **8**(2), 157–175 (1997)
4. Feige, U.: A threshold of $ln\ n$ for approximating set cover. J. ACM **45**(4), 634–652 (1998)
5. Feld, S.: The focused organization of social ties. Am. J. Sociol. **86**(5), 1015–1035 (1981)
6. Friedkin, N.E.: Horizons of observability and limits of informal control in organizations. Soc. Forces **62**(1), 54–77 (1983)
7. Granovetter, M.S.: The strength of weak ties. Soc. Netw. **1977**, 347–367 (1977)
8. Huang, H., et al.: Chapter 5 Analysis and Prediction of Triadic Closure in Online Social Networks. Social Network Analysis, pp. 105–136 (2017)
9. Krackhardt, D.: Cognitive social structures. Soc. Netw. **9**(2), 109–134 (1987)
10. Kuhn, H.W.: The Hungarian method for the assignment problem. In: 50 Years of Integer Programming 1958-2008, pp. 29–47 (2009)
11. Leskovec, J., et al. (2008). Microscopic evolution of social networks. In: Proceeding of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD 08 (2008)
12. Lewis, K.: Measuring transactive memory systems in the field: scale development and validation. J. Appl. Psychol. **88**(4), 587–604 (2003)
13. Li, M., et al.: A coevolving model based on preferential triadic closure for social media networks. Sci. Report. **3**(1), 2013 (2013)
14. March, J.G.: Exploration and exploitation in organizational learning. Organ. Sci. **2**(1), 71–87 (1991)
15. Mollenhorst, G., et al.: Shared contexts and triadic closure in core discussion networks. Soc. Netw. **33**(4), 292–302 (2011)
16. Moreland, R.L., et al. Socially shared cognition at work: transactive memory and group performance. In: What's Social about Social Cognition? Research on Socially Shared Cognition in Small Groups, pp. 57–84 (1996)
17. Nemhauser, G.L., et al. (1978). An analysis of approximations for maximizing submodular set functions–I. In: Mathematical Programming, vol. 14, no. 1, pp. 265–294 (1978)
18. Puterman, M.L.: Chapter 8 Markov decision processes. In: Handbooks in Operations Research and Management Science, vol. 2, pp. 331–434 (1990)
19. Ren, Y., et al.: Transactive Memory Systems 1985–2010: An Integrative Framework of Key Dimensions, Antecedents, and Consequences. The Academy of Management Annals **5**(1), 189–229 (2011)

20. Snijders, T.A.B., et al.: Introduction to stochastic actor-based models for network dynamics. Soc. Netw. **32**(1), 44–60 (2010)
21. Wegner, D.M.: Transactive memory: a contemporary analysis of the group mind. In: Mullen, B., Goethals, G.R. (eds.) Theories of Group Behavior, pp. 185–208. Springer, New York (1987)