2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)

Togetherness: An Algorithmic Approach to Network Integration

Anastasia Moskvina
Auckland University of Technology
Private Bag 92006, Auckland, New Zealand
Telephone: (64) 99219999, Fax: (64) 99219944
anastasia.moskvina@aut.ac.nz

Jiamou Liu
The University of Auckland
Private Bag 92019, Auckland, New Zealand
Telephone: (64) 9 923 9528
jiamou.liu@auckland.ac.nz

***Abstract*—Network integration refers to a process of building links between two networks so that they dissolve into a single unified network. *Togetherness* measures the proximity of these two networks as they integrate; this notion is fundamental to social networks as it is relevant to important concepts such as trust, coherence and solidarity. In this paper, we study the algorithmic nature of network integration and formally introduce three notions of togetherness. We analyze the corresponding computational problems of network integration: Given two networks and a desired level of togetherness, build links between members of these networks so that the overall network meets the togetherness criterion. We analyze optimal solutions to this problem, describe several heuristics and compare their performance through experimental analysis.**

***Keywords*—Network integration, togetherness, distance, collaboration networks**

I. Introduction

The establishment of links has been a fundamental question in the study of complex networks. While links often emerge due to natural network evolution, there are many scenarios where links are created "by design", i.e. connections are set up in order to meet certain targets. Take, as an example, the design of flight routes of airlines to ensure effective service, or, the intermarriages between the Medicis with other XV century noble European families for the purpose of gaining power in Florence [20]. Forging new links between two disjoint networks brings these networks together. Several questions naturally arise concerning such processes: How do two departments in an organization merge into a single unit? How do two research teams socialize and collaborate? How to bridge existing bus routes to create a unified public transport map? How to create hyperlinks connecting two web domains to allow convenient browsing?

Motivated by the questions above, we address the algorithmic nature of network integration. The problem asks to build links between members of two networks so that the combined network becomes a unified whole. It is then a major question how "together" the unified networks should be as an outcome of this process. Naturally, the more links there are that connect the two networks, the closer they become. On the other hand, there is normally a cost associated with establishing and maintaining links. Therefore, it is important to strike a balance between the amount of *togetherness* and the number of new links created between the networks. We now describe the three main goals of the paper:

The **first goal** concerns over the notion of togetherness. In recent years there has been a surge of the use of "togetherness" in sociology [14], communication studies [3], politics [22] and biology [6]. In its most original form, togetherness is a concept rooted in Kantian philosophy, meaning the confluence of intuition and concepts [9]. In mathematics, togetherness is regarded as a "mark of being integrated into a single unity" and influences the creations of notions such as continuity and connectedness [13]. The notion is first discussed in information science by cybernetic pioneer Gordon Pask in his 1980 essay [21]; Pask refers togetherness as an "index of human proximity" that is "determined by a communication/computation medium". He goes on to discuss how togetherness can be "engineered" through a process of "conversation", which is abstractly represented as the integration of two concept networks. In this paper, we rediscover and follow this seminal work, and provide a formal interpretation of togetherness in the context of network science.

The **second goal** of the paper deals with different levels of togetherness. Already in the work of Pask, it is mentioned that an appropriate measure of togetherness comes from the notion of *distance*. In a network, the distance between two nodes is the smallest number of "hops" needed to move from one node to the other. It is natural to adopt distance as an indication of togetherness. In particular, *diameter* refers to the largest distance between nodes in the network. It is well-known that most real-world networks enjoy small diameters – this is the so-called *small world* property. We hold the view that all nodes of a network have certain resources; and when a network has a small diameter, the resource on each node can be reached out from everyone else within a few steps, and each member is able to influence others. Hence, the diameter of the integrated network forms the strongest form of togetherness.

When expressing togetherness between two networks in their integration, diameter may be too strong. We define further two weaker notions of togetherness. Firstly, *existential togetherness* considers distances between every node in one network to some node in the other network. Secondly, *universal togetherness* considers distances between every node in one network to all nodes in the other network. The former measure of togetherness may be reasonable if we assume all nodes in

IEEE/ACM ASONAM 2016
August 18-21, 2016, San Francisco, CA, USA
978-1-5090-2846-7/16/$31.00

any network hold the same resource, and it is enough to reach any node in a network. The latter measure of togetherness may be reasonable if the distances to all nodes in the other network are important. In this paper, we relate and compare these three notions of togetherness.

The **third goal** of the paper concerns with the algorithmic problem of *network integration*: Given two networks G_1, G_2 and a desired value for a specific type of togetherness, we would like to compute a small set of links to be set up between G_1, G_2 so that the integrated network meets the togetherness requirement. We study computational complexity of this problem and propose methods that generate solutions. The first type of methods are heuristics that are based on the equi-privilege properties of the networks. The second type are simulations based on certain priorities given to nodes of the network. To compare these methods,we perform experimental analysis on both synthesized and real-world data.

Related Works: This work can be regarded as a continuation of our previous study [15] that explores algorithms for socializing a newcomer into an established social network. It also extends results presented in [16], where we studied how new links could be established between two social groups assuming that all persons have the same privilege.

The work is relevant to the following interrelated research areas: Firstly, *strategic network formation* considers how new links emerge due to rational and self-centred decisions of members of the network [10]. This field puts focus on stability and equilibrium that arise out of game-theoretical situations, but does not emphasis network integration, nor does it concern with togetherness of the resulting network. Secondly, *interdependent networks* discusses the complex structures formed through an integration of networks of different types (e.g. a transportation network with an electrical network) [7]; the focus here is mainly on interdependence among the nodes and robustness of network, i.e. whether node failures leads to a cascading failure throughout the overall infrastructure. Lastly, we mention management studies on *collaborative team building*. When two companies merge, the success of the new entity largely hinges on whether the firms can effectively socialize employees from both sides to a unified direction [1]. The challenge lies in how venues could be set up (e.g. meetings, group assignments, etc.) that nurture collaboration and allow efficient communication [23], [24]. The framework proposed in this paper addresses this challenge through an algorithmic perspective.

Paper organization.: Section II presents the main definitions which include three togetherness notions for network integration. Section III discusses the network integration problems, and analyze their optimal solutions. Section IV proposes algorithmic mechanisms for solving the network integration problems. Section V presents experimental results. Finally, Section VI concludes with remarks on future works.

II. Main Definitions

A. Preliminaries

We define a *network* as a connected undirected unweighted graph $G = (V, E)$ where V is a set of *nodes* and E is a set of (undirected) *edges* on V. Let $G = (V, E)$ be a network. We write an edge $\{u, v\} \in E$ as uv and say that u, v are *adjacent*. A *path* (of *length* k) is a sequence of nodes $u_0, u_1, \ldots, u_k$ where $u_i u_{i+1} \in E$ for any $0 \leq i < k$. The *distance* between u and v, denoted by $\mathsf{d}(u, v)$, is the length of a shortest path from u to v. The *eccentricity* of u is $\mathsf{ecc}(u) := \max_{v \in V} \mathsf{d}(u, v)$. The *diameter* of the network G is $\mathsf{diam}(G) := \max_{u \in V} \mathsf{ecc}(u)$. The *radius* of G is $\mathsf{r}(G) := \min_{v \in V} \mathsf{ecc}(v)$. A node $u \in V$ is called a *center* of G if $\mathsf{ecc}(u) = \mathsf{r}(G)$.

A set $D \subseteq V$ is a *dominating set* for G if every node not in D is adjacent to at least one member of D. The *dominating number* $\gamma(G)$ is the number of nodes in a smallest dominating set for G. More generally, for $\ell \in \mathbb{N}$, a set $D_\ell \subseteq V$ is a *distance-ℓ dominating set* if every node not in D is at distance at most ℓ from $u \in D$. The *ℓ-dominating number* is the number of nodes in a smallest distance-ℓ dominating set for G.

For two disjoint sets of nodes V_1, V_2, we use $V_1 \otimes V_2$ to denote the set of all edges $\{uv \mid u \in V_1, v \in V_2\}$; these edges will be our instruments for integration two networks with nodes V_1 and V_2, respectively.

Definition 1: Let $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ be two networks. Fix a non-empty set of edges $E \subseteq V_1 \otimes V_2$. The *integrated network* of G_1 and G_2 by E is $G_1 \oplus_E G_2 := (V_1 \cup V_2, E_1 \cup E_2 \cup E)$.

B. Togetherness

Consider the integration $G_1 \oplus_E G_2$. Any edge $uv \in E$ represents a channel for the flow of certain resources (information, traffic, knowledge, etc.) between G_1 and G_2. Hence, the set E should provide nodes in each network with appropriate access to resources in the other network. *Togetherness* is an index for proximity of G_1 and G_2, and thus measures the effectiveness of E. As argued above, distances between nodes play a significant role. Further we introduce three levels of togetherness and motivate each notion with an example scenario in organizational management:

(a): Imagine two groups of specialists who provide information and advices to each other (e.g. the accounting and the procurement teams in a company). A member of one group needs to access some but not necessarily all members of the other group. In this case, it is sufficient to measure togetherness based on the distance from a node in a network to any node in the other network. In particular, the *$\exists$-span* $\sigma_E^{\exists}(u)$ of $u \in V_i$ refers to $\min\{\mathsf{d}(u, v) \mid v \in V_{3-i}\}$ where $i \in \{1, 2\}$; let $\sigma_E^{\exists}(G_1, G_2) := \max\{\sigma_\exists(u) \mid u \in V_1 \cup V_2\}$.

(b): Imagine two groups of people with varying skills who collaborate on a joint project. To fully utilize skills and incorporate knowledge, a person in one group should access everyone in the other group. Hence we measure togetherness based on the distance from a node in a network to all members of the other network. In particular, the *$\forall$-span* $\sigma_E^{\forall}(u)$ of $u \in V_i$ refers to $\max\{\mathsf{d}(u, v) \mid v \in V_{3-i}\}$ where $i \in \{1, 2\}$; let $\sigma_E^{\forall}(G_1, G_2) := \max\{\sigma_\forall(u) \mid u \in V_1 \cup V_2\}$.

(c): Imagine two groups of people who merge into a single group. To ensure the resulting group is a cohesive, tightly-knit unit, we measure togetherness based on the diameter of the combined group.

Definition 2: Let $G_1 \oplus_E G_2$ be an integration of two networks $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$. We define three notions of togetherness of G_1 and G_2 as follows:

1) The $\exists$-*togetherness* (or *existential togetherness*) is defined as $\tau_E^{\exists}(G_1, G_2) \coloneqq \left(\sigma_E^{\exists}(G_1, G_2)\right)^{-1}$.
2) The $\forall$-*togetherness* (or *universal togetherness*) is defined as $\tau_E^{\forall}(G_1, G_2) \coloneqq \left(\sigma_E^{\forall}(G_1, G_2)\right)^{-1}$.
3) The Δ-*togetherness* (or *diametric togetherness*) is defined as $\tau_E^{\Delta}(G_1, G_2) \coloneqq (\mathsf{diam}\,(G_1 \oplus_E G_2))^{-1}$.

When G_1, G_2 and E are clear from context, we abuse the notation writing $\chi^{\natural}$ for $\chi_E^{\natural}(G_1, G_2)$ for all $\chi \in \{\sigma, \tau\}$ and $\natural \in \{\exists, \forall, \Delta\}$. In the following we use d and $\widetilde{d}$ to denote $\max\{\mathsf{diam}(G_1), \mathsf{diam}(G_2)\}$ and $\min\{\mathsf{diam}(G_1), \mathsf{diam}(G_2)\}$, respectively.

Proposition 3: The following properties hold for all networks $G_1 = (V_1, E_1), G_2 = (V_2, E_2)$ and $E \subseteq V_1 \otimes V_2$:

(a) $\left(\sigma^{\exists} + \widetilde{d}\right)^{-1} \leq \tau^{\forall} \leq \tau^{\exists}$

(b) $\tau^{\Delta} \leq \tau^{\forall}$; and $\tau^{\forall} = \tau^{\Delta}$ whenever $\sigma^{\forall} \geq d$

Proof: For (a), it is clear that $\tau^{\forall} \leq \tau^{\exists}$ as $\sigma^{\exists}(u) \leq \sigma^{\forall}(u)$ for every node u. Without loss of generality assume $\mathsf{diam}(G_2) \leq \mathsf{diam}(G_1)$. From any node $u \in V_1$, there is $v \in V_2$ where $\mathsf{d}(u, v) \leq \sigma^{\exists}$, and for all $w \in V_2$, $\mathsf{d}(v, w) \leq \widetilde{d}$. Thus $\mathsf{d}(u, w) \leq \tau^{\exists} + \widetilde{d}$. This means that $\sigma^{\forall} \leq \sigma^{\exists} + \widetilde{d}$ and hence $\tau^{\forall} \geq \left(\sigma^{\exists} + \widetilde{d}\right)^{-1}$.

For (b), it is clear that $\tau^{\Delta} \leq \tau^{\forall}$ as $\mathsf{diam}\,(G_1 \oplus_E G_2) \leq \sigma^{\forall}(u)$ for any node u. When $\sigma^{\forall} \geq d$, $\mathsf{d}(u, v) \geq \max\{\mathsf{diam}(G_1), \mathsf{diam}(G_2)\}$ for any $u \in V_1$ and $v \in V_2$. Thus $\mathsf{diam}(G_1 \oplus_E G_2) = \sigma^{\forall}$, which means $\tau^{\forall} = \tau^{\Delta}$. ■

As an example, we integrate two networks in three ways in Fig. 1.

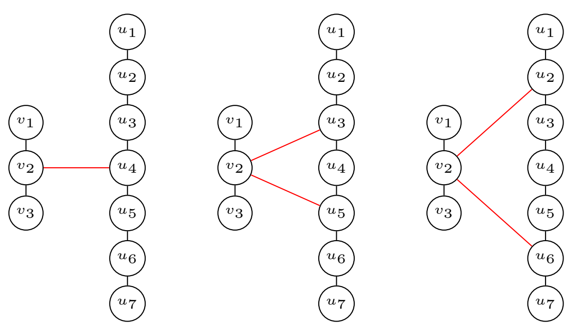


Fig. 1: Integrating two line networks: $E_1 = \{v_2u_4\}$ with $\tau^{\exists} = 1/4$, $\tau^{\forall} = 1/5$, and $\tau^{\Delta} = 1/6$ (on the left); $E_2 = \{v_2u_3, v_2u_5\}$ with $\tau^{\exists} = 1/3, \tau^{\forall} = 1/4$, and $\tau^{\Delta} = 1/6$ (in the middle); and $E_3 = \{v_2u_2, v_2u_6\}$ with $\tau^{\exists} = 1/3$, $\tau^{\forall} = \tau^{\Delta} = 1/4$ (on the right).

III. The Network Integration Problems

When integrating two networks G_1 and G_2, we have two constraints: The first constraint is the togetherness of G_1 and G_2 in the integrated network. The second constraint is the number of new edges established during the process. To ensure high togetherness, one needs to create sufficiently many edges between G_1 and G_2. As each edge requires certain resources to set up and maintain, the challenge is to allow maximal togetherness while creating minimal number of new edges. Formally, fix $\natural \in \{\exists, \forall, \Delta\}$. We define the following problems:

1) *Network Integration under Togetherness constraint* $\mathsf{NIT}_t^{\natural}(G_1, G_2)$ (where $t \in (0, 1]$): This problem asks for, given $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$, a set of edges $E \subseteq V_1 \otimes V_2$ such that the togetherness $\tau_E^{\natural}(G_1, G_2) \geq t$. An *optimal solution* E of this problem is one that has the smallest cardinality.
2) *Network Integration under Edge constraint* $\mathsf{NIE}_e^{\natural}(G_1, G_2)$ (where $e \geq 1$ is an integer): This problem asks for, given $G_1 = (V_2, E_1)$ and $G_2 = (V_2, E_2)$, a set $E \subseteq V_1 \otimes V_2$ that has cardinality e. An *optimal solution* E of this problem is one that leads to the largest togetherness $\tau_E^{\natural}(G_1, G_2)$.

These two problems are closely related.

Theorem 4 (Duality): For any $\natural \in \{\exists, \forall, \Delta\}$, there is a solution of $\mathsf{NIT}_t^{\natural}(G_1, G_2)$ containing at most e edges iff there is a solution E of $\mathsf{NIE}_e^{\natural}(G_1, G_2)$ that leads to $\tau^{\natural} \geq t$.

The next result discusses the size of any optimal solution E of $\mathsf{NIT}_1^{\natural}(G_1, G_2)$ (i.e., when we desire maximal togetherness).

Lemma 5: For any networks G_1, G_2 and $E \subseteq V_1 \otimes V_2$:

(a) $\tau^{\exists}{=}1$ iff $\forall u \in V_i \exists v \in V_{3-i} \colon uv \in E$ where $i \in \{1, 2\}$.

(b) $\tau^{\forall}{=}1$ iff $E = V_1 \otimes V_2$.

(c) $\tau^{\Delta}{=}1$ iff both G_1, G_2 are complete and $E = V_1 \otimes V_2$.

Proof: For (a), let $E \subseteq V_1 \otimes V_2$ be a set of edges with cardinality $|E| = \max\{|V_1|, |V_2|\}$. If E connects every node in V_1 with some node in V_2 and vice versa, then the $\exists$-span $\sigma^{\exists}(u) = 1$ for every $u \in V_1 \cup V_2$. On the other hand, if $\exists u \in V_1 \forall v \in V_2 \colon uv \notin E$, then $\sigma^{\forall}(u) \geq 2$.

For (b), it is sufficient to note that if there is some $u \in V_1, v \in V_2$ with $uv \notin E$, then $\mathsf{d}(u, v) \geq 2$; (c) is straightforward as $\tau^{\Delta} = 1$ iff $G_1 \oplus_E G_2$ is a complete graph. ■

The next result discusses togetherness achieved by optimal solutions of $\mathsf{NIE}_1^{\natural}(G_1, G_2)$ (i.e., adding one new edge $\{u, v\}$).

Lemma 6: For any networks G_1, G_2 and $uv \in V_1 \otimes V_2$, the maximum value of $\tau_{\{uv\}}^{\natural}(G_1, G_2)$ is

(a) $(\max\{\mathsf{r}(G_1), \mathsf{r}(G_2)\}{+}1)^{-1}$ if $\natural = \exists$.

(b) $(\mathsf{r}(G_1){+}\mathsf{r}(G_2) + 1)^{-1}$ if $\natural = \forall$.

(c) $(\max\{\mathsf{r}(G_1){+}\mathsf{r}(G_2){+}1, d\})^{-1}$ if $\natural = \Delta$.

Proof: The optimal solution $\{uv\}$ connects a center u in G_1 with a center v in G_2. The properties (a)(b)(c) can then be easily checked. ■

The problem of finding optimal solutions for network integration is in general computationally hard.

Theorem 7: For any $\natural \in \{\exists, \forall, \Delta\}$, the following problems are hard for $\mathsf{W}[2]$, the second level of the W-hierarchy:

1) Fix $t \in (0, 1/2]$. Decide if $\mathsf{NIT}^\natural_t(G_1, G_2)$ has a solution with $\le e$ edges for given G_1, G_2 and integer $e > 0$
2) Fix $e > 1$. Decide if $\mathsf{NIE}^\natural_e(G_1, G_2)$ has a solution E that leads to $\tau^\natural \ge t$ for given G_1, G_2 and $t \in (0, 1]$.

Proof: Due to the duality (Thm. 4) between the two problems $\mathsf{NIT}^\natural_t(G_1, G_2)$ and $\mathsf{NIE}^\natural_e(G_1, G_2)$, it is sufficient to prove one of (1) and (2). As shown in [12], finding the smallest distance-r dominating set in G with diameter $r+1$ is complete for $\mathsf{W}[2]$ (for any fixed r). We now show a reduction from this problem to $\mathsf{NIT}^\natural_t(G_1, G_2)$ for $t \in (0, 1/2]$.

Suppose, without loss of generality, that $t = k^{-1}$ for some integer $k \ge 2$. Now let $G_1 = (V_1, E_1)$ be a graph with diameter k and let G_2 be a graph that contains only a single node $\{u\}$. For any distance-$(k-1)$ dominating set $S \subseteq V_1$, the set of edges $S \otimes \{u\}$ is a solution of $\mathsf{NIT}^\natural_t(G_1, G_2)$. Conversely, suppose $S \subseteq V_1$ is not distance-$(k-1)$ dominating. Then there is a node $w \in V_1$ that is at distance at least k away from any node $v \in S$. This means that $\mathsf{d}(w, u)$ in the integrated network is at least $k+1$ and S is not a solution of $\mathsf{NIT}^\natural_t(G_1, G_2)$. Thus $\mathsf{NIT}^\natural_t(G_1, G_2)$ has a size-ℓ solution if and only if G_1 has a size-ℓ distance-$(k-1)$ dominating set. ■

In subsequent sections we focus on heuristics for solving the $\mathsf{NIT}^\natural_t(G_1, G_2)$ and $\mathsf{NIE}^\natural_e(G_1, G_2)$ problems. We first focus on $\exists$-togetherness and characterize the optimal solutions of $\mathsf{NIT}^\exists_t(G_1, G_2)$.

Theorem 8 ($\exists$-Togetherness Theorem): Suppose E is an optimal solution of $\mathsf{NIT}^\exists_t(G_1, G_2)$. Then

(1) If $t = 1$, then $|E| = \max\{|V_1|, |V_2|\}$

(2) If $t < (\max\{\mathsf{r}(G_1), \mathsf{r}(G_2)\})^{-1}$, then $|E| = 1$

(3) If $1 > t \ge (\max\{\mathsf{r}(G_1), \mathsf{r}(G_2)\})^{-1}$, $|E| = \max\{\gamma_1, \gamma_2\}$, where γ_i is the $(t^{-1}-1)$-dominating number of the network G_i for each $i \in \{1, 2\}$.

Proof: (1) and (2) directly follow from Lem. 5(a) and Lem. 6(a), resp. We now prove (3).

Let $k = t^{-1}$ and $D_1 \subseteq V_1$ and $D_2 \subseteq V_2$ be minimum distance-$(k-1)$ dominating sets for G_1 and G_2, resp. In other words, $|D_1| = \gamma_1$ and $|D_2| = \gamma_2$. Without loss of generality, assume $\gamma_1 \ge \gamma_2$. Then there is a set $E \subseteq V_1 \otimes V_2$ that contains for every $u \in D_i$, some edge uv where $v \in D_{3-i}$ where $i \in \{1, 2\}$, and $|E| = \gamma_1$. Our goal is to show that E is an optimal solution of $\mathsf{NIT}^\exists_t(G_1, G_2)$.

Note that any node w in V_i is at most $k-1$ steps away from some node $u \in D_i$, which means that the $\exists$-span $\sigma^\exists(w) \le k$. Thus E is a solution of $\mathsf{NIT}^\exists_t(G_1, G_2)$. Now take a set $E' \subseteq V_1 \otimes V_2$ has $|E'| < \Gamma_1$. Let $S = \{u \in V_1 \mid \exists v \in V_2 \colon uv \in E'\}$. Then there is some node $w \in V_1$ that is at least k steps away from any node in S. Thus the $\exists$-span $\sigma^\exists(w) > k$ and E' is not a solution. This means that E is an optimal solution. ■

For $\forall$- and Δ-togetherness, the next theorem bounds the size of optimal solutions of $\mathsf{NIT}^\natural_t(G_1, G_2)$ for large t (i.e., $t \ge 1/3$). Recall that $\gamma(G)$ denotes the dominating number of G.

Theorem 9: Suppose E be an optimal solution of $\mathsf{NIT}^\natural_t(G_1, G_2)$ where $\natural \in \{\forall, \Delta\}$.

(1) If $t = 1$, then $|E| = |V_1| \cdot |V_2|$

(2) If $t = 1/2$, then $|E| \le \min\{\gamma(G_1) \cdot |V_2|, \gamma(G_2) \cdot |V_1|\}$

(3) If $t = 1/3$, then $|E| \le |V_1| + |V_2| - 1$

Proof: (1) directly follows from Lem. 5 (b)(c). For (2), let D_1, D_2 be a dominating set in G_1 and G_2, respectively. Let $E_1 = \{uv \mid u \in D_1, v \in V_2\}$ (so $|E_1| = \gamma(G_1) \cdot |V_2|$) and $E_2 = \{uv \mid u \in V_1, v \in D_2\}$ (so $|E_2| = \gamma(G_2) \cdot |V_1|$). Then both $G_1 \oplus_{E_1} G_2$ and $G_1 \oplus_{E_2} G_2$ have diameter 2, and thus $\tau^\forall = \tau^\Delta = 1/2$.

For (3), pick any node $u \in V_1$ and $v \in V_2$ and let $E' = \{uy \mid y \in V_2\} \cup \{xv \mid x \in V_1\}$. Then $|E'| = |V_1| + |V_2| - 1$. The diameter of the integrated network $G_1 \oplus_{E'} G_2$ is 3, and thus $\tau^\forall = \tau^\Delta = 1/3$. ■

IV. Methods for Network Integration

We present several algorithms for integrating two networks. The mechanisms are broadly divided into two categories: 1) We propose two heuristics that search for small sets E that integrate two networks G_1, G_2. We assume that the networks enjoy *equi-privilege property*, i.e., any pair of nodes between networks can be freely connected. The goal of these heuristics is to gain maximal togetherness in the integrated networks. 2) We propose four scenarios where nodes in one network preferentially establish links with nodes in the other network. Here every node is given a *priority* which is determined by the network structure. The difference between these priority-based methods and the heuristics in the first category is that their aim is to simulate the preferential attachments of links during integration, rather than explicitly searching for good solutions.

A. Integrating Networks with Equi-privilege Property

Discussions on *equi-privilege property* originates from organizational behavioral studies of social networks. In [4], Mila Baker describes peer-to-peer organizations as social structures where members have equal privileges regardless of their roles (such as volunteer groups, research teams, etc.); these organizational structures are said to have equi-privilege property. The main challenge of merging two such organizations is to establish channels that allow exchanges of intellectual ideas and trusted transactions.

The mechanism for network integration depends on the level of togetherness one desires to achieve. If the goal is to optimize $\exists$-togetherness, by Theorem 8, the key is to identify dominating sets in the networks G_1 and G_2. If the goal is to optimize $\forall$- or Δ- togetherness, then it is desirable to establish links that minimize diameter.

Optimizing $\exists$-togetherness ($\mathsf{MinLeaf}$)**:** Man and Duckworth in [8] propose a heuristic to produce small distance ℓ-dominating sets for a given ℓ in *regular graphs*, i.e., graphs where all nodes have the same degree. We modify this heuristic to an algorithm $\mathsf{MinLeaf}(G_1, G_2, t)$ that integrates networks with $\exists$-togetherness constraint. Our algorithm takes two networks $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ and $t \in (0, 1]$ as input and outputs a set $E = D_1 \oplus D_2$, where D_1 and D_2 are distance-$(t^{-1}-1)$ dominating sets in G_1 and G_2, respectively. By Theorem 8, the set E is a solution of $\mathsf{NIT}^\exists_t(G_1, G_2)$.

For $i \in \{1,2\}$, the algorithm iteratively builds $D_i \subseteq V_i$ by maintaining a set $U_i \subseteq V_i$ of "uncovered" nodes, i.e., nodes that have distance $\geq k$ from any current node in D_i. The initial configuration is when $U_i = V_i$. It repeatedly performs the following operations until $U = \varnothing$:

1) Select a node $u \in U$ and add u to D_i (see below).
2) Compute all nodes at distance at most $(k-1)$ from v and remove these nodes from U.

We now describe the heuristic for finding a node u in each iteration. A node is called a *leaf* if it has minimum degree in the graph; leaves correspond to least connected members in the network, and may become outliers once nodes with higher degrees are removed from the network. Therefore, the heuristic first picks a leaf v in U_i, then applies a sub-procedure to find the next node u to be added to U_i. The sub-procedure determines a path $v = w_1, w_2, \dots$ iteratively as follows:

1) Suppose w_i is picked. If $i = r$ or w_i has no adjacent node in U_i, set w_i as u and terminate the process.
2) Otherwise select a w_{i+1} (which is different from w_{i-1}) among adjacent nodes of w_i with maximum degree.

When this process terminates, the algorithm adds u to D_i. Note that the distance between u and v is at most $k-1$.

Optimizing $\forall$-togetherness (CtrPer)**:** We now propose another heuristic to solve the $\mathsf{NIT}_t^{\beta}(G_1, G_2)$ problem for $\beta \in \{\forall, \Delta\}$. When integrating G_1 and G_2, it makes sense first to establish a link between centres of the networks, as they have the closest proximity to other nodes. Then, if x, y are nodes that are furthest apart in the integrated network, they share the "weakest channel", represents a form of *structural hole*. Hence, it makes sense to connect x, y by an edge. Formally, the *center* $C(G)$ of a graph $G = (V, E)$ is the set of all nodes that have the least eccentricity, i.e., $C(G) = \{v \in V \mid \mathsf{ecc}(v) = \mathsf{r}(G)\}$. A pair of nodes (x, y) in G forms a *peripheral pair*, denoted by $(x, y) \in \mathsf{P}(G)$, if $\mathsf{d}(x, y) = \mathsf{diam}(G)$.

The $\mathsf{CtrPer}(G_1, G_2, t)$ algorithm takes as input networks G_1, G_2 and $t \in (0, 1]$, and iteratively builds a sequence of edges $E = \{e_0, e_1, \dots\}$ such that (1) $e_0 \in C(G_1) \otimes C(G_2)$; and (2) for all $i \geq 1$, $e_i \in \mathsf{P}\left(G_1 \oplus_{\{e_0, e_1, \dots, e_{i-1}\}} G_2\right)$. The process stops whenever the $\forall$-span $\sigma^{\forall}(u)$ of every node $u \in V_1 \cup V_2$ is no more than t^{-1}.

Theorem 10: For any networks G_1, G_2 and $t \in (0, 1]$, the $\mathsf{MinLeaf}$ algorithm outputs a solution of the $\mathsf{NIT}_t^{\exists}(G_1, G_2)$ problem, the CtrPer algorithm outputs outputs a solution of the $\mathsf{NIT}_t^{\forall}(G_1, G_2)$ problem.

B. Integrating Networks with Priority-based Approaches

In real-world networks nodes differ in various ways which may affect their integration. For example, in an organization, people sometimes connect to each other according to their abilities or roles. We therefore consider the network integration problem under the assumption that all nodes have certain priorities. We investigate cases when higher priorities are assigned to nodes with different structural properties:

1) $\mathsf{MaxDegree}$: A high degree indicates the possession of certain advantage such as capability or resources. Hence, we give higher priorities to nodes with higher degrees.

2) $\mathsf{MinDegree}$: A low degree indicates a certain disadvantage such as isolation and lack of resources. To ensure togetherness, it also may be reasonable to give higher priorities to nodes with lower degree.

3) MaxBtw: *Betweenness* indicates the centrality of a node, i.e., how much the node serve as a "gatekeeper" and connects diverse parts of the network [5]. Hence in this scenario, we give higher priorities to nodes with higher betweenness.

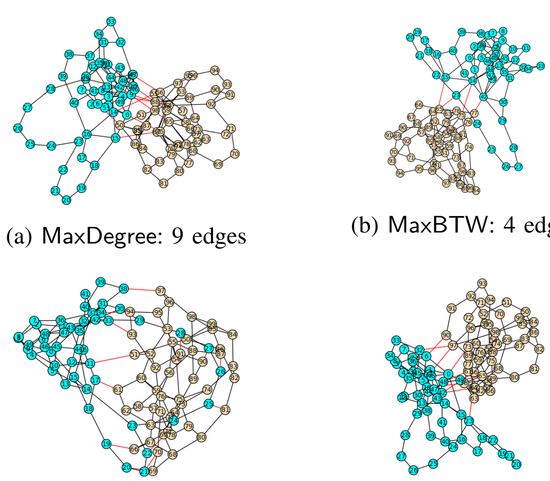
4) Random: Lastly, we consider the case when the priorities are assigned randomly. This corresponds to a case when the priorities are assigned according to some extraneous factors.

For each of the four scenarios above, we implement a mechanism that integrates networks G_1 and G_2 to achieve $\natural$-togetherness $t \in (0, 1]$. The procedure iteratively builds a set $E \subseteq V_1 \otimes V_2$ that is a solution to $\mathsf{NIT}_t^{\natural}(G_1, G_2)$. If uv is created, the nodes u, v become *bridging nodes* that link G_1 with G_2. We have the following intuition:

1) To allow smooth flow of resources between the two networks and avoid *information gate keepers*, we should have many different bridging nodes.
2) Nodes with higher priorities should serve more as bridging nodes and be linked.

Suppose a set of edges E' has already been created. We adopt the following mechanism to find two nodes $u \in V_1$ and $v \in V_2$: For any node $w \in V_i$, let $b(w) = |\{v \in V_{3-i} \mid wv \in E'\}|$. Choose the node u from $B_1 = \{w \in V_1 \mid b(w) \leq b(w') \text{ for all } w' \in V_1\}$ that has the highest priority; Choose the node v from $B_2 = \{w \in V_2 \mid b(w) \leq b(w') \text{ for all } w' \in V_2\}$.

To illustrate the ideas above, Fig. 2 shows the result of integrating two Newman-Watt-Strogatz random networks (see Sec V) with 50 nodes each using various approaches, and arrive at an integrated network with diameter 9. The result shows that, while MaxBtw requires the smallest number of edges, there is a big variation in terms of the number of edges created using different priorities. It is therefore interesting to compare the results of the different heuristics in more detail.



(a) $\mathsf{MaxDegree}$: 9 edges

(b) MaxBTW: 4 edges

(c) $\mathsf{MinDegree}$: 14 edges

(d) Random: 10 edges

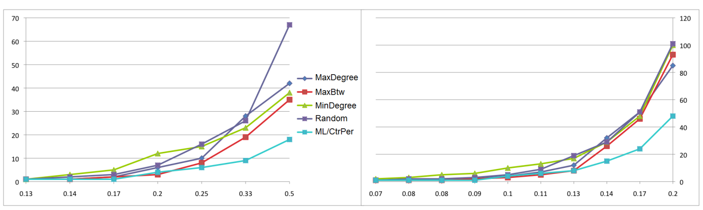
Fig. 2: Integrating two Newman-Watts-Strogatts networks with 50 nodes to achieve diameter of 9 in the integrated network.

V. Experimental Analysis

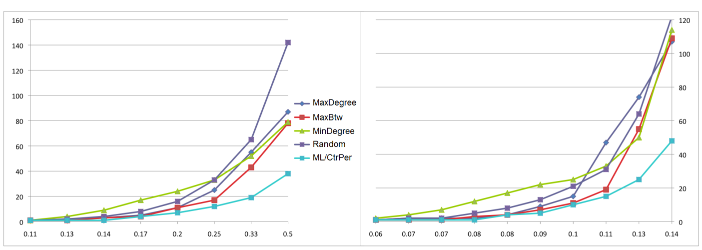
We implement all heuristics mentioned above and compare their results. For our experiments, we generate two types of random graphs. The first (NWS) is Newman-Watts-Strogatz's small-world network model [19], which produces graphs with small average path lengths and high clustering coefficient. The second (BA) is Barabasi-Albert's preferential attachment model, which generates scale-free graphs whose degree distribution of nodes follows a power law; this is an essential property of numerous real-world networks [2].

A. Experiment 1. Solving $\mathsf{NIT}_t^{\natural}(G_1, G_2)$

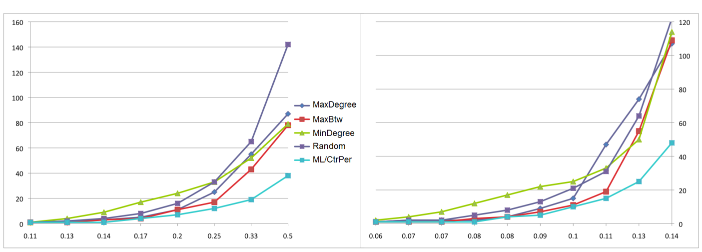
We generated 10 pairs of NWS and BA networks with 50, 100 and 200 nodes each. For each pair we compute a solution for the $\mathsf{NIT}_t^{\natural}(G_1, G_2)$ problem (where $\natural \in \{\exists, \forall\}$) using MaxDegree, MinDegree, MaxBtw, Random as well as MinLeaf (when $\natural = \exists$) and CtrPer (when $\natural = \forall$). Fig. 3 and Fig. 4 display the average number of new edges in the solution sets for the NWS and BA networks, resp. The number of edges increases with increasing $\tau^{\exists}$ and $\tau^{\forall}$. Furthermore, for small togetherness ($\tau^{\exists} \leq 0.17$ for NWS and $\tau^{\exists} \leq 0.25$ for BA, and $\tau^{\forall} > max\{d(G_1), d(G_2)\}^{-1}$), different types priorities do not significantly affect the size of the resulting sets. However, the difference increases as togetherness increases; in general, the MinLeaf and CtrPer algorithms output much smaller edge sets.



(a) Networks with 50 nodes

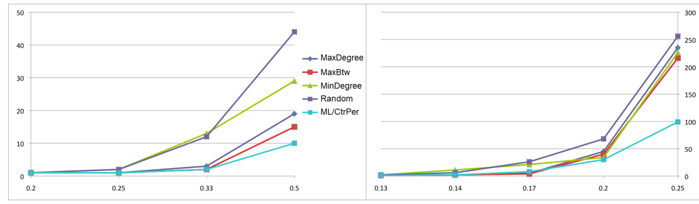


(b) Networks with 100 nodes

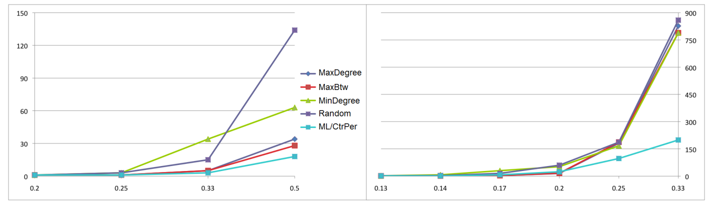


(c) Networks with 200 nodes

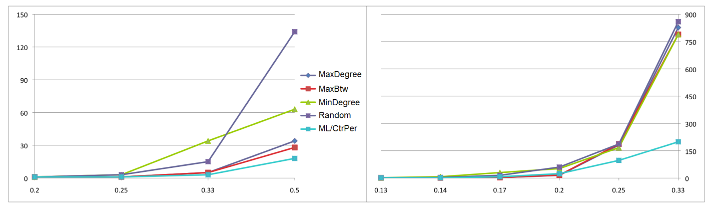
Fig. 3: Comparing heuristics: average numbers of edges required to integrate two NWS networks with fixed $\tau^{\exists}$ (on the left) and fixed $\tau^{\forall}$ (on the right)



(a) Networks with 50 nodes



(b) Networks with 100 nodes



(c) Networks with 200 nodes

Fig. 4: Comparing heuristics: average numbers of edges required to integrate two BA networks with fixed $\tau^{\exists}$ (on the left) and fixed $\tau^{\forall}$ (on the right)

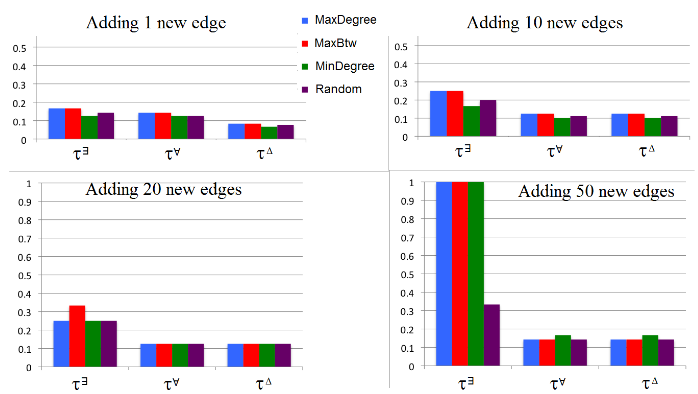


Fig. 5: Integrating NWS networks by establishing 1, 10, 20 and 50 edges

B. Experiment 2. Solving $\mathsf{NIE}_e^{\natural}(G_1, G_2)$

We generate networks with 50 nodes and set the numbers of edges e to values $1, 10, 20, 50$. We compute the average togetherness in the integrated networks by applying different heuristics. Fig. 5 and Fig.6 plot the results for NWS and BA networks, resp. The best performance is given by MaxBtw. In general, MinDegree gives the worst performance when e is small. However, its performance catches up with other heuristics when e becomes larger. On the contrary, Random has an opposite behavior: togetherness grows slower as more

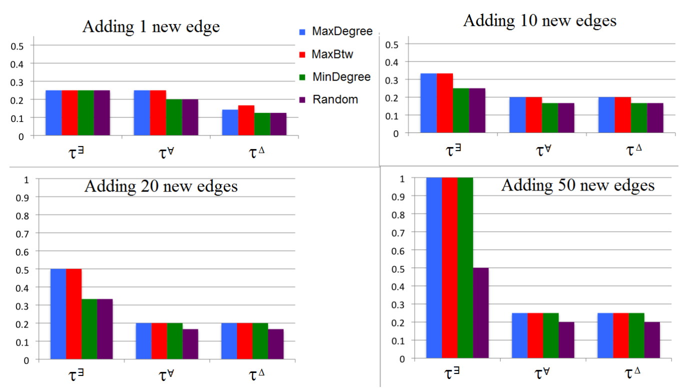
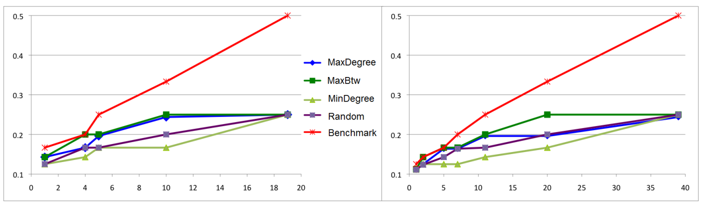
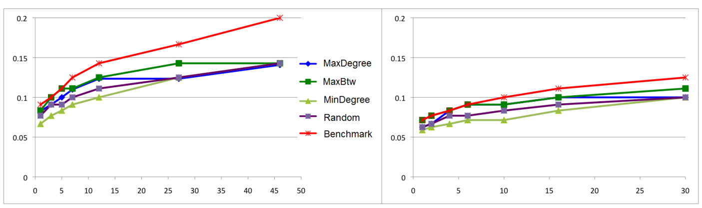


Fig. 6: Integrating BA networks by establishing 1, 10, 20 and 50 edges



(a) Average values of $\tau^{\exists}$ for NWS networks



(b) Average values of $\tau^{\forall}$ for NWS networks

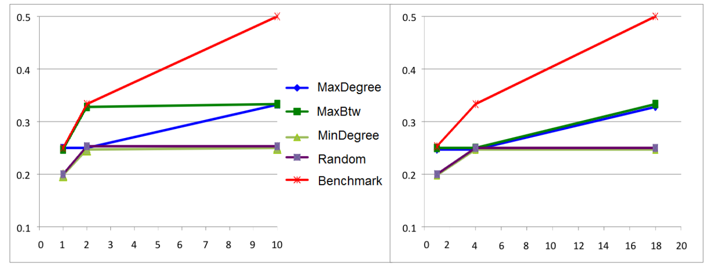
Fig. 7: Integrating NWS networks with 50 nodes (on the left) and 100 nodes (on the right)

edges are randomly added.

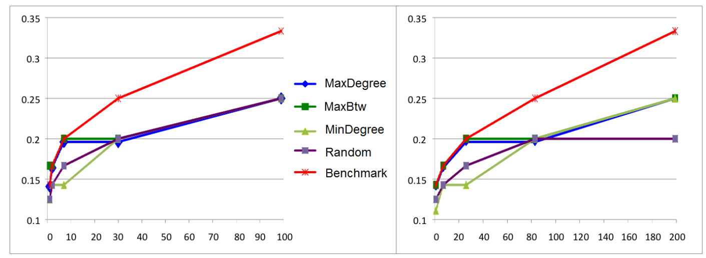
C. Experiment 3. Priority-based methods

As the MinLeaf and CtrPer algorithm in general give small solution sets for the integration problems, we first apply them and use the resulting solution size as *benchmarks* to test the performance of the priority-based methods. For each value of $\exists$- and $\forall$-togetherness, we calculate the average number e of edges in the output solution sets. Then, we apply the priority-based heuristics to compare the result of these methods against the benchmarks. The resulting togetherness (as well as the benchmarks) are plotted in Fig. 7 and Fig. 8.

The results show that, when we add a small number of edges, the priority-based heuristics perform well: the MaxBtw method results in the same togetherness as the benchmark. The MinDegree method, as in the previous experiment, proves to be the worst for small number of edges, however performs better when more edges are added. Rather surprisingly, integrating networks with the random strategy often produce solutions that are comparable with the other strategy.



(a) Average values of $\tau^{\exists}$ for BA networks



(b) Average values of $\tau^{\forall}$ for BA networks

Fig. 8: Integrating BA networks with 50 nodes (on the left) and 100 nodes (on the right)

	Collaboration1	Collaboration 2
total number of nodes	5,242	9,877
total number of edges	14,496	25,998
number of nodes in the subgraph	4,158	8,638
number of edges in the subgraph	13,422	24,806
diameter	17	18
radius	9	10

TABLE I: Collaboration 1 and Collaboration 2 datasets

D. Real World Datasets

We use two real datasets to reconfirm the results obtained for the synthesized datasets. Col1 and Col2 are networks that represent scientific collaborations in General Relativity, Quantum Cosmology (Col1), and in High Energy Physics Theory (Col2) [11]. Table I lists details of these networks. As both networks are initially unconnected, we considered the giant component in each graph.

We first apply the MinLeaf and CtrPer algorithms. Structurally Col1 and Col2 resembles the small-world networks of NWS models. Thus, we expect that when integrating these two networks, despite the large number of nodes and edges, the number of new edges would be relatively small. In Fig. 9, we show how the togetherness measures $\tau^{\exists}$ and $\tau^{\forall}$ change with the number of added edges. The weaker togetherness notion, $\tau^{\exists}$, grows faster than $\tau^{\forall}$: decreasing the $\forall$-togetherness is harder. Consistent with our expectation, we notice that very few edges are required when integrating these large collaboration networks.

We then apply the priority-based methods to the networks Col1 and Col2. Similarly to Exp. 3, we fix the number of added edges according the benchmarks provided by MinLeaf and CtrPer and then apply the different priority-based methods. The results are plotted in Figure 10, which shows that, in

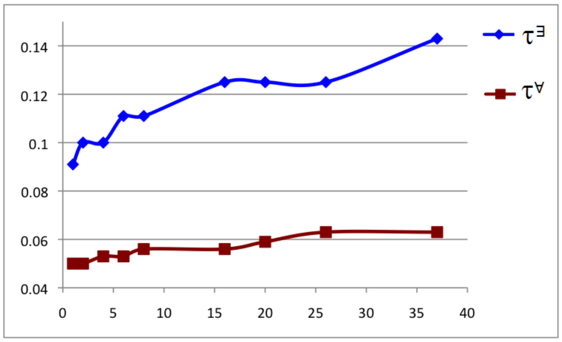


Fig. 9: Integrating two collaboration networks with $\tau^{\exists}$ and $\tau^{\forall}$ constraints

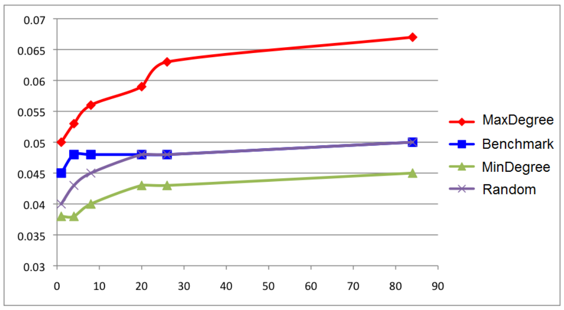


Fig. 10: Integrating two collaboration networks: comparing different strategies

general, the Random method gives comparable performance against other priority-based strategies.

VI. Conclusion and Future Works

Network integration amounts to the fundamental question that arises in numerous social, political and physical domains. Our earlier work in [15] focuses on how an individual algorithmically establishes links to socialize into a network. Here, we extend this effort to explore the integration of two arbitrary networks. The novelty lies in that we apply a formal framework and employ various heuristics to tackle the problem. The key conceptual contribution is in proposing three measures of togetherness, which are useful indication of proximity between sub-networks. We believe that togetherness will be helpful not only in this context, but in any problem domains where solidarity and distances are of concern.

Contrary to intuition, our experiments demonstrate that the random strategy for building links performs comparable to other heuristics in a few situations; It would be an interesting future work to explore the mathematical reason behind this phenomenon, e.g., what is the expected togetherness if we connect two random graphs using k random edges.

It would also be interesting to incorporate node characteristics in surrounding contexts and apply other principles, e.g., homophily, to guide the establishment of links. Another future work is to incorporate directed or weighted edges in the networks. A potential application is to develop technology that advise potential links or collaborations (say, in an online social platform) to members of two social groups.

Another direction of future work is to investigate network building in the context of organizational networks by incorporating both formal and information relations, the model we presented in [17].

References

[1] R. Aguilera, J. Dencker, and Z. Yalabik, *Institutions and Organizational Socialization: integrating employees in cross-border mergers and acquisitions*. International Journal of Human Resource Management, 15(8): 1357–1372, 2006

[2] R. Albert, and A.-L. Barabási, *Statistical mechanics of complex networks.* Reviews of Modern Physics 74, 47–97, 2002.

[3] M. Bakardjieva, *Virtual togetherness: An everyday-life perspective*. Media, Culture and Society, 25, 291–313, 2003

[4] M., Baker, *Peer-to-Peer Leadership: Why the Network Is the Leader*, Berrett-Koehler Publishers, 2013.

[5] M. Barthélemy, *Betweenness centrality in large complex networks*. Eur. Phys. J. B 38, 163–168, 2004.

[6] E. Cox, and J. Bonner, *The Advantages of Togetherness*, Science, 292 (5516), 448–449, 2001.

[7] M. Danziger, A. Bashan, Y. Berezin, L. Shekhtman, and S. Havlin, S., *An Introduction to Interdependent Networks*, Nonlinear Dyn. of Elec. Sys., 189–202, 2014.

[8] W. Duckworth and B. Mans. *Randomized Greedy Algorithms for Finding Small k-Dominating Sets of Regular Graphs.* Random Structures and Algorithms 10/2005; 27(3):401 – 412, 2005.

[9] R. Hanna, *Kant and Non-conceptual Content*, European Journal of Philosophy 13: 247-90, 2005.

[10] M. Jackson, *A Survey of Models of Network Formation: Stability and Efficiency*, Group Formation in Economics, Cambridge University Press, 11–57, 2005.

[11] J. Leskovec, J. Kleinberg and C. Faloutsos, *Graph Evolution: Densification and Shrinking Diameters.* ACM TKDD, 1(1), 2007.

[12] D. Lokshtanov, N. Misra, N., G. Philip and M. Ramanujan, M. *Hardness of* r*-dominating set on graphs of diameter* $(r-1)$, in Proc. of IPEC 2013, 255–267, 2013.

[13] J.R. Lucas, *Conceptual Roots of Mathematics*, International Library of Philosophy, Routledge, 2011

[14] J.R. McCarthy, *The powerful relational language of family: Togetherness, belonging and personhood*. The Sociological Review, 60(1), 68–90. 2012.

[15] A. Moskvina, J. Liu, *How to Build Your Networks? A Structural Analysis*, IJCAI 2016, to appear.

[16] A. Moskvina, A., J. Liu, *Integrating Networks of Equipotent Nodes*, In Proc. of CsoNet2016, to appear.

[17] A. Moskvina, J. Liu, *Hierarchies, ties and power in organizational networks: Model and analysis*, In ASONAM Proc. of the 2015 IEEE/ACM Intern. Conf. on Advances in SNAM, pp. 202–209, 2015.

[18] M. Newman, *Mathematics of networks*, in The New Palgrave Encyclopedia of Economics, 2nd edition, L. E. Blume and S. N. Durlauf (eds.), Palgrave Macmillan, Basingstoke (2008).

[19] M. Newman, D. Watts and S. Strogatz, *Random graph models of social networks*. Proceedings of the National Academy of Science of USA, 99:2566–2572, 2002.

[20] J. Padgett, and C. Ansell, *Robust Action and the Rise of the Medici, 1400-1434*, American Journal of Sociology, 98(6):1259–1319, 1993.

[21] G. Pask, *The Limits of Togetherness*, in S.H. Lavington (ed.) Proc. Information Processing'80. North-Holland, Amsterdam (1980).

[22] K. Strøm and W. Müller, *The keys to togetherness:Coalition agreements in parliamentary democracies, The Journal of Legislative Studies*, 5:3-4, 255–282, 1999.

[23] T.E. Stuart and O. Sorenson *Strategic networks and entrepreneurial ventures*, Strategic Entrepreneurship Journal 1 (3-4), 211-227, 2007

[24] T.E. Stuart, *Network positions and propensities to collaborate: An investigation of strategic alliance formation in a high-technology industry*, Administrative science quarterly, vol 43, No 3 (Sep., 1998), pp 668-698.