# Web and Society: A First Look into the Network of Human Service Providers

Charalampos Chelmis
*Department of Computer Science*
*University at Albany, SUNY*
Albany, NY, USA
cchelmis@albany.edu

Mengfan Yao
*Department of Computer Science*
*University at Albany, SUNY*
Albany, NY, USA
myao@albany.edu

Wonhyung Lee
*School of Social Welfare*
*University at Albany, SUNY*
Albany, NY, USA
wlee@albany.edu

***Abstract*—Human service organizations (HSOs) operate in an environment considered to be prohibitive of collaboration. To understand how HSOs come together to address the grand challenges associated with meeting human needs, we attempted to automatically construct the network of HSOs based on the information publicly available through each organization's website–the medium that people use to find relevant information to access services. Our analysis of the the complex system of relationships among HSOs in Albany, New York suggests that the network of HSOs in this area exhibits a multipolar structure with few super connectors, and strong relations between organizations that serve similar functions. We quantitatively evaluate the quality of the constructed HSOs' network from Web data based on structured, in–person interviews we conducted with HSOs.**

***Index Terms*—Computational social science, applied network science**

## I. Introduction

Human service organizations (HSOs) [1] operating within a region are often faced with the dual pressure to compete as well as coordinate their operations to enhance delivery of human services (e.g., food, shelter) [2]–[5]. Because of this dilemma, it has long been suspected that HSOs may be operating in a relatively self–contained manner, leading to a potentially redundant, confusing, or incoherent configuration of services [6]–[9]. Nevertheless, critical insights into how HSOs are connected to one another beyond administrative coordination are limited [5], [10]. A network of HSOs can be constructed from interview data, where nodes represent organizations and edges denote partnerships between such organizations. However, interview data may be difficult if not prohibitive to obtain due to reasons including, but not limited to, (i) resource constraints (e.g., unavailability of HSOs' staff for interviews), (ii) low HSO response rate [11], and (iii) scalability (i.e., the time required to collect data from face–to–face in–depth interviews prohibits using the same method for a large number of HSOs).

To address this challenge, we take a data–driven computational social science approach, by using the Internet to collect data about how HSOs connect to one another. Specifically, we systematically collect semi–structured information from HSOs' websites, including their mission and scope and partner organizations, into a directed and attributed network, where nodes denote HSOs and edges represent partnerships between them. We focus on HSOs in the metropolitan area surrounding Albany, the capital of the New York state. Albany, the 4th largest metropolitan region in the state, and the 45th largest in the U.S., is a multiracial and multiethnic city that contains nearly $99,000$ residents (on average $52\%$ White, $29\%$ Black or African American, $9\%$ Hispanic or Latino, $6\%$ Asian, and growing refugee populations)[1] and a high concentration of human service providers and service organizations related to government, health care, and education [12].

The advantage of using websites to collect information about HSOs is twofold. First, they are easy to collect; websites are publicly available online and their content is machine–readable. Second, they provide rich and semi–structured data in the form of explicit connections between documents, that can be used to translate the data directly into a network, which can be subsequently analyzed to provide insights into the interactions among HSOs. The disadvantage of using Web data is that it may lead to an imperfect representation of the actual network of partnerships between HSOs, since HSOs may choose not to list their partners on their websites or they may use offline resources such as printed books similar to yellow pages to connect to other organizations.

To evaluate the quality of the constructed network from Web data (as opposed to interview data), we performed a small–scale evaluation based on structured, in–person interviews of 43 HSOs in Albany, New York between April and October 2018. Our results indicate that although many HSOs may not disclose their actual partners on the Web, information that is available can indeed be collected with very high accuracy.

This material is based upon work supported by the National Science Foundation under Grant No. ECCS–1737443.



*ASONAM '19*, August 27–30, 2019, Vancouver, BC, Canada

ACM ISBN 978-1-4503-6868-1/19/08...$15.00
http://dx.doi.org/10.1145/3341161.3345331

[1]http://www.census.gov/programs-surveys/acs/technical-documentation/table-and-geography-changes/2014/5-year.html

Therefore, we argue that the capacity of collecting and analyzing data about how HSOs interact with an unprecedented breadth, depth, and scale has the potential to unlock new perspectives on inter–organizational networks.

Our analysis has identified several organizations that play a central role in connecting HSOs in the local service coordination efforts and suggests that the network of HSOs in Albany exhibits a multipolar structure, instead of the typical strong core–periphery structure often found in complex networks [13], with almost equally structurally important communities. Based on these findings, organizations can target the key connectors and influencers to initiate service coordination or bring a new coordination model in the region. Future research can further investigate the characteristics of isolated network communities and whether it would be worthwhile to facilitate service coordination with other organizations in the region.

## II. Related Work

Networks among organizations have long been studied [14], [15] in various organizational environments, including finance, natural resource management, and healthcare (e.g., [2], [16]–[19]), offering critical insights into how organizations relate to one another on a basis of power and influence, as well as how organizations create collaborations to manage competition or resource interdependencies.

Nevertheless, little is known about the organizational network of human services. While network analysis has been used with an eye towards strengthening community partnerships, and service coordination and integration [3], only a few studies have examined organizational collaboration outcomes based on direct surveys with the organizations (e.g., [5], [10]), and not from the lens of the general public who attempts to navigate services through such organizations. Our study presented in this paper thus fills a unique gap in the literature as it investigates the network of human service organizations by analyzing publicly available, real–life data.

## III. Methodology

The data for this analysis was collected during June 2018 based on a seed list of HSOs and their URLs, which are provided as input to a web crawler implemented in Python. The seed list was compiled from (i) GuideStar,[2] a service specializing in reporting on U.S. nonprofit companies, (ii) GreatNonprofits,[3] a website that allows donors, volunteers, and clients to share their personal experiences with and reviews of charitable organizations, essentially providing crowdsourced information about the reputability of these organizations, and (iii) Charity Navigator,[4] an independent charity watchdog that evaluates charitable organizations in the United States.

By human service organizations (HSOs), we refer to human services and community improvement organizations as defined by the National Taxonomy of Exempt Entities Classification

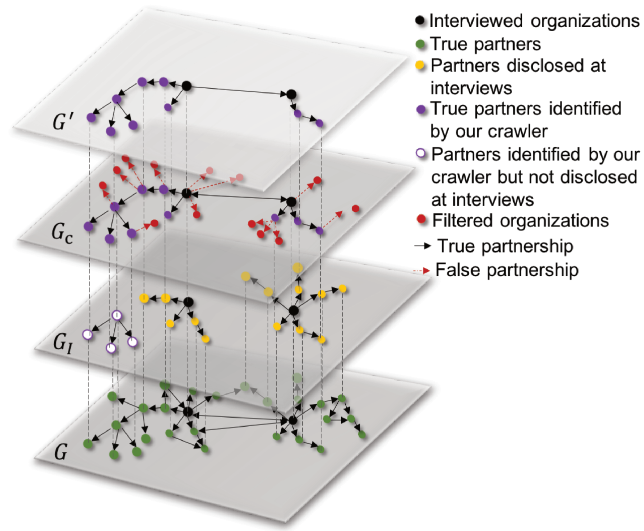


Fig. 1: Illustration (better in color) of the true, yet unobserved, partnership network $\mathcal{G}$, and networks $\mathcal{G}_\mathcal{I}$ and $\mathcal{G}_C$, obtained manually from interviews and automatically from the Web, respectively. Irrelevant organizations in $\mathcal{G}_C$ are excluded by human experts resulting in the refined network $\mathcal{G}' \subseteq \mathcal{G}_C$.

System (NTEECS)[5]. Specifically, HSOs correspond to all major groups I, J, K, L, M, N, O, and P under category "V. Human Services", and major group S under category "VII. Public, Societal Benefit".

The starting domain of a given website was examined for partner organizations (i.e., organizations listed in a "partners" page) while links to social media platforms such as Twitter were filtered so that no websites outside the realm of human service organizations are considered. We additionally impose a depth limit of 3 (i.e., discard webpages that are more than 3 clicks away from the home page), meaning that each domain may have potentially not been fully crawled. Our web crawl performs snowball sampling, i.e., crawling organizations that are listed as partners on websites of organizations in our seed list, to further expand the original list to a total of over $3,000$ service–providing organizations spanning multiple domains, not necessarily HSOs. After excluding organizations that do not operate in Albany, their main focus is not human service delivery (i.e., major hospitals and large medical centers, government agencies, faith–based organizations, school districts, membership–based associations and self–help groups such as Alcoholics Anonymous, and programs of colleges and universities), or may be defunct[6], a total of $848$ nodes, spanning 23 categories[7], remain (see Table I).

[2]https://www.guidestar.org/

[3]https://greatnonprofits.org/

[4]https://www.charitynavigator.org/

[5]https://nccs.urban.org/classification/national-taxonomy-exempt-entities

[6]We annotated an organization as defunct if we found its website to be inaccessible after a total of 4 trials over our data collection period.

[7]As recorded by the National Taxonomy of Exempt Entities system.

### A. Challenges

We came across a number of challenges related to data originating human service organizations' websites:

- We use the URL of an organization to identify it as unique. In some cases however, different URLs may resolve to the same webpage (e.g., due to parameters appended to the URL which do not change the content of the page). On the other side, some organizations may use the same domain to host their webpages.
- Some organizations have minimalistic websites containing a general overview of their activities, while other organizations might have multiple pages for specific programs and services, and may list partners differently under each program, or may opt to not list partners altogether. The same is true regarding information about the mission of organizations.
- Partnering organizations might appear in various formats, such as in menu items, images, and PDF files.
- Automatic categorization of organizations is a challenging problem in itself.

To address the above challenges, we (i) performed post–analysis to identify URLs resolving to the same organization and (ii) cross–referenced information from GuideStar, GreatNonprofits, and Charity Navigator to assign a category to organizations already in our database. We manually assigned categories to organizations beyond our initial list.

With respect to identifying webpages listing partners, an easy solution would be to automatically filter navigation menus to identify pages that contain information about partner organizations. However, given differences and lack of standards in web design, this may be infeasible when crawling a number of heterogeneous websites. While cognizant of such difficulties, we have instructed our web crawler to traverse a given domain looking for web pages containing at least three URLs with a ".org" suffix, as indicative of human service organizations, and therefore a potentially good heuristic to identify webpages that list partners. In future work, we plan to experiment with additional heuristics in order to enhance the ability of our web crawler to parse additional formats to better capture existing relationships between organizations.

### B. Network Representation

Figure 2 demonstrates how we map crawled data into a directed partnership graph. Specifically, a bipartite graph of organizations and webpages, on which they are listed as partners is created as shown on the left. The bipartite graph is then transformed to a directed graph where an edge is drawn from the organization making its partners publicly available to each organization listed as its partner. Formally, we represent the partnership network as a directed graph $G = (V, E)$, where the set of nodes $V = \{v_1, \ldots, v_N\}$ represent organizations, and set $E = \{e_{ij} | i, j \in V\}$ denotes edges among organizations as discovered by our web crawler; an edge $e_{ij}$ from node $i$ points to node $j$ if organization $j$ is listed as a partner on $i$'s website. We have chosen this representation to capture the non–reciprocal nature of information availability on organizations' websites (i.e., organization $i$ listing $j$ as a partner but not vise versa).

| Id | Category | # of nodes |
|---|---|---|
| 1 | Arts, Culture & Humanities | 10 |
| 2 | Education | 69 |
| 3 | Environment | 10 |
| 4 | Health Care | 83 |
| 5 | Mental Health & Crisis Intervention | 73 |
| 6 | Voluntary Health Associations & Medical Disciplines | 52 |
| 7 | Medical Research | 13 |
| 8 | Crime & Legal-Related | 35 |
| 9 | Employment | 19 |
| 10 | Food, Agriculture & Nutrition | 27 |
| 11 | Housing & Shelter | 40 |
| 12 | Recreation & Sports | 18 |
| 13 | Youth Development | 20 |
| 14 | Human Services | 112 |
| 15 | International, Foreign Affairs & National Security | 6 |
| 16 | Civil Rights, Social Action & Advocacy | 15 |
| 17 | Community Improvement & Capacity Building | 154 |
| 18 | Philanthropy, Voluntarism & Grant making Foundations | 16 |
| 19 | Science & Technology | 9 |
| 20 | Social Science | 2 |
| 21 | Public & Societal Benefit | 31 |
| 22 | Religion–Related | 32 |
| 23 | Mutual & Membership Benefit | 2 |

TABLE I: Distribution of nonprofit organizations operating in Albany across categories.

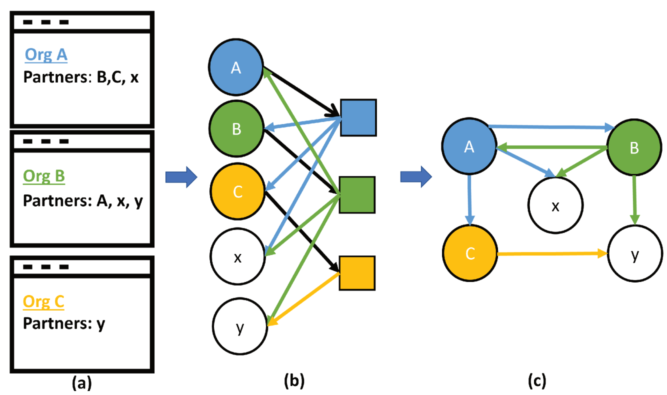


Fig. 2: We map webpages listing partnership information (a) into a bipartite graph (b), where circles denote organizations and squares encode webpages. The bipartite graph is subsequently projected into a directed graph (c). Although not all organizations have a web presence (e.g., $x$ and $y$ in this example) they may still be identified as partners if they appear in at least one webpage.

### C. Quality of Network Construction from Web Data

To quantify our web crawler's ability to collect relevant data, we performed a small–scale evaluation as follows. Let $\mathcal{G}$ be the true, unobserved network of HSOs and their partners in the real–world, and $\mathcal{G}_{\mathcal{I}}$ be the network of HSOs and the partners each HSO disclosed to us during the interviews. Additionally, let $\mathcal{G}_C$ denote the network constructed by our crawler with the interviewed HSOs as the seed list. Finally, let $\mathcal{G}' \subseteq \mathcal{G}_C$ be the refined constructed network obtained after human experts exclude irrelevant organizations, such as medical centers, from $\mathcal{G}_C$. Note that organizations not mentioned during the interviews can still be discovered online. We denote such "surprise" nodes as $V_s = \{v|v \in V_{\mathcal{G}'} - V_{\mathcal{G}_{\mathcal{I}}}\}$. An illustrative example of these networks is shown in Figure 1. For reference, the sizes of $\mathcal{G}_{\mathcal{I}}$, $\mathcal{G}_C$ and $\mathcal{G}'$ are 256, 503, and 75 respectively, while the size of $\mathcal{G}$ is unknown.

Next, we evaluate the quality of our crawler, as well as the availability of partnership information on providers' websites. Specifically, we compute the precision and recall of our crawler by comparing $\mathcal{G}_C$ to $\mathcal{G}'$. With a recall of 92%, our crawler is able to discover partnerships in a given website quite accurately. However, this comes to the expense of efficiency, quantified by a low precision (15%), indicating that in its current state, our crawler cannot effectively discriminate between irrelevant organizations and HSOs. This is due to the fact that it only relies on url suffix and the address of organizations to exclude potentially irrelevant organizations and organizations operating outside of our study site. In our current work, we are exploring machine learning methods to improve the precision of our crawler. Finally, to evaluate the correspondence of partnership information listed online to the real–world, we rely on $\mathcal{G}_{\mathcal{I}}$ as a coarse approximation of $\mathcal{G}$, and use it as the ground truth. Among the interviewed organizations, only 5 list partners on their website, and among all listed partners, only 7 organizations are also mentioned as partners during the interviews. This results in a recall of 7%, implying that most organizations do not disclose their actual partners on their website. At the same time, 55% precision indicates that only a small fraction of partners is actually disclosed during the interview process.

### D. High Level Statistics

Here, we compare the properties of the network of HSOs (i.e., $\mathcal{G}'$) with other technological and biological networks in the literature. We find the ratio of the total number of edges to the number of nodes to be $\sim 1$, even though the distribution of the number of partnerships differs widely among organizations. This confirms the long suspected hypothesis that service providing organizations operate in silos [6]–[9]. Similarly, the degree of symmetry in the network is only 0.33%. Although the percentage of symmetric links in the largest connected component increases slightly to 0.37%, it is significantly lower than values reported for other complex networks [20].

We now look at clustering, which quantifies the degree of how densely the neighborhood of a node is connected. Not all nodes are connected in one cluster. Instead, the network of human service providers comprises 8 connected components, the largest of which encompasses 96.4% of the network. The global density of the network, measured by its clustering coefficient, is $c = 0.013$. In a random network with the same number of nodes (i.e., $|V|$) and degree $d$, $c_{rand} = \frac{d}{|V|} = 0.0025$. Next, we examine the properties of shortest paths between organizations in the large weakly connected component (WCC). We found the average path length and diameter to be 5.35 and 13, respectively. For comparison, in a random network with the same number of nodes and average degree, the expected diameter is $\frac{\ln(d)}{\ln(|V|)} = 1.03$, whereas the average path length on the Web, if it were to be treated as an undirected graph, is 7 [21]. We additionally computed a variance of relative diameter for each category $k$, by considering the subgraph composed of nodes $V_k \subseteq V$ in $k$, and using as edge length between two nodes $i, j \in V_k$ the shortest path between such nodes (potentially through nodes in $V \setminus V_k$). The inset in Figure 3 illustrates the computation of this relative diameter per category on a toy network with nodes into two categories, indicated by blue circles and orange squares, respectively. Note that by definition the relative diameter for each category is less than the diameter of the network, as even the longest shortest path between any two nodes in the network is less than the network diameter. The reason for using this atypical definition of relative diameter is to study the connectivity between organizations embedded in the larger network. Simply filtering nodes by category would result in a disconnected subgraph of $G$, and thus ignor potential pathways between nodes in $V_k$ through nodes in $V \setminus V_k$. In sociology, this is known as the boundary specification problem [22].

Figure 3 shows how the relative diameter varies across categories. Intuitively, categories represented by only few nodes, such as 15 (International, Foreign Affairs & National Security), 20 (Social Science), and 23 (Mutual & Membership Benefit) tend to exhibit smaller diameters compared to larger categories. A large diameter may be indicative of larger paths, however, it may also be interpreted as a way of identifying categories whose nodes are dispersed across the network (i.e., categories with high diameter) as opposed to categories whose nodes may be forming closely connected groups.

## IV. SUPER–CONNECTORS AND ELITES

While the network of HSOs resembles social and technological networks, there are significant differences. Most notably, the network is not a small–world since the average path length between nodes in the giant component and the network diameter are large (c.f. Section III-D). By comparison, the network of Fortune 800 firms in the 1970s, although of similar size, was shown to have a much smaller diameter [23]. The existence of long paths indicates limited coordination of HSOs.

A small number of nodes of very high degree connects a large number of organizations to the rest of the network. Such super–connectors can be empirically confirmed by testing for a heavy–tailed degree distribution. To test how well the in– and out–degree distributions are modeled by power–law, we calculated the best power–law fit using the maximum likelihood method [24], which minimizes the Kolmogorov–Smirnov statistic, $D$, between the cumulative distribution function of the data and the power law: $\hat{\alpha} = arg\max_a D$, where $D = max_x|P_{emp}(x) - P_a(x)|$, and $P_{emp}(x)$ and $P_a(x)$

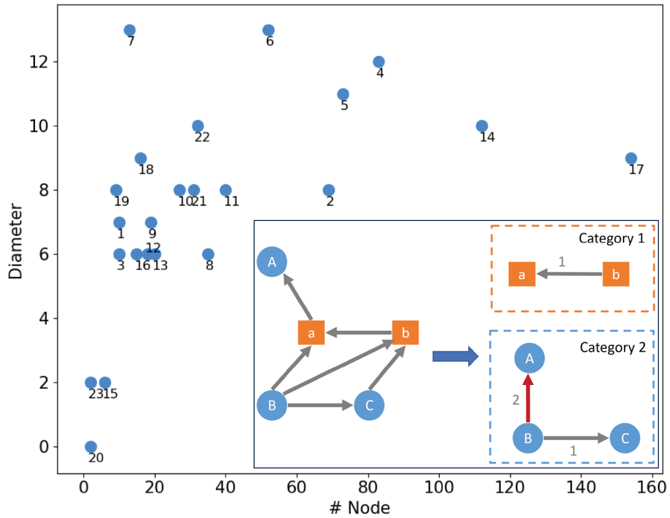


Fig. 3: Relative diameter (y–axis) over category size, i.e., number of nodes (x–axis) for all 23 categories. The inset shows a toy example where nodes $a$ and $b$ belong to category 1, and nodes $A$, $B$ and $C$ belong to category 2. The length of path from node $B$ to $A$ (red arrow) is 2 in category 2, as the shortest path between $A$ and $B$ is through node $a$.

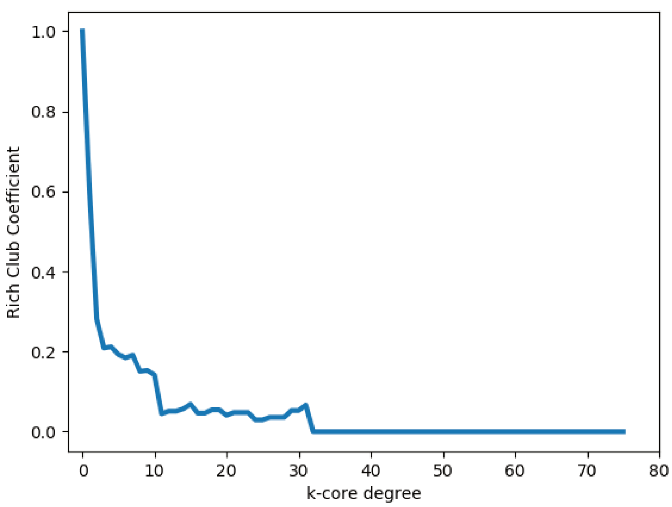


Fig. 4: Rich–club coefficient as a function of $k-$core. High degree nodes (i.e., network elites) connect almost exclusively to non–elite nodes.

denote the cdfs of the data and the power law with exponent $a$, respectively. We found the estimated power–law coefficient to be $1.45, 2.59$, and $2.19$ for the the in–, out– and total degree distributions respectively, with a Kolmogorov–Smirnov goodness of fit of $0.19, 0.12$, and $0.1$ accordingly, suggesting that power–law approximates the distributions very well. This finding is consistent with power laws in complex networks [25], and has profound implications with respect to search over the network of HSOs [26], and its fragility [27].

To identify nodes important both for the connectivity of the network as well as information flow over it (i.e., nodes that create shortcuts, and hubs that connect otherwise distant nodes) we compute closeness, eigenvector, and betweeness centrality [28]–[30]. Table II shows the top 5 nodes ranked by their centrality score. The New York state government[8] and the Troy Rehabilitation and Improvement Program[9] rank consistently high across these metrics indicating their prevalent position in the network. Catholic Charities of the Diocese of Albany,[10] which has a high closeness centrality score, is one of the largest, private, social service agencies in the Capital Region that help people across fourteen counties address basic human needs such as food, clothing, and shelter, and provide support for people with developmental disabilities, pregnant and parenting teens, domestic violence victims, and the homeless. UnitedWay of the Greater Capital Region,[11] which exhibits a high betweeness centrality score, brings together nonprofits, businesses, government and human service agencies, schools, organized labor, financial institutions, community development corporations, voluntary and neighborhood associations, and the faith community to address basic needs, education, income and health issues in the Capital Region.

To characterize the connectivity of super–connectors, we tested for the rich–club phenomenon [31] as shown in Figure 4. Evidently, the rich–club coefficient drops sharply as a function of $k-$core indicating that nodes with higher degree are far more likely to connect to low–degree nodes. This finding comes at odds with classic results for corporate networks, where co–located firms are far more likely to form alliances [17], despite the significant density of HSOs in Albany [12].

Finally, to understand interactions within and across categories (see Table I), we consider an aggregate network in which nodes are categories and edges denote the cumulative number of links between nodes that belong to different categories, as shown in Figure 5. The size of nodes is proportional to the number of organizations of a given category, and edge weights denote the number of partnerships stemming out from a given category. We find that within category partnerships (indicated by a thick self–loop edge) are more common for categories smaller in size (e.g., category 8). Instead, more inter–category ties are observed among categories that serve similar functions or share the same purpose. Consider for example the strong connectivity of category 17 (Community Improvement & Capacity Building) with 11 (Housing & Shelter), and category 9 (Employment) and 14 (Human Service).

## V. The Multipolar Structure of Networked Human Services

To study the large–scale structure of the network of HSOs, we partition $\mathcal{G}'$ into communities using a bottom–up, multi–level clustering algorithm, where each node is moved to other communities iteratively so as to maximize the local contribution to the total modularity score [32]. The network of

[8] https://www.ny.gov/
[9] http://www.triponline.org/
[10] http://www.ccrcda.org/
[11] https://www.unitedwaygcr.org/

| Closeness Centrality | Betweeness Centrality | Eigenvector Centrality |
|---|---|---|
| Government of New York State | United Way | Niskayuna Central School |
| Mediation Matters Organization | Regional Food Bank | Port of Albany |
| Catholic Charities | Hunger Solutions | Government of New York State |
| Troy Rehabilitation and Improvement Program | Center for Child | Troy Rehabilitation and Improvement Program |
| Equinox | Niskayuna Central School | Albany Community Land Trust |

TABLE II: Top 5 organizations sorted by their structural importance in the network of HSOs.

| Group | Size | Average In–/Out–Degree | Major Category & Coverage | PageRank |
|---|---|---|---|---|
| 1 | 100 | 1.13±0.4/ 1.28±5.26 | 2 (21.74%) | 0.001167 ± 0.0022 |
| 2 | 29 | 0.99±0.2/ 1.04±8.62 | 19 (85.9%) | 0.001165 ± 0.0044 |
| 3 | 87 | 1.48±1.04/ 1.4±5.5 | 16 (23.75%) | 0.001246 ± 0.0022 |
| 4 | 44 | 0.98±0.15/ 1.0±6.63 | 5 (52.27%) | 0.001178 ± 0.0035 |
| 5 | 45 | 1.25±1.7/ 1.0±3.44 | 6 (41.67%) | 0.001205 ± 0.0017 |
| 6 | 36 | 0.83±0.41/ 0.83±2.04 | 5 (66.67%) | 0.001179 ± 0.0011 |
| 7 | 6 | 0.99±0.08/ 1.01±12.77 | 16 (31.45%) | 0.001173 ± 0.0068 |
| 8 | 159 | 1.32±0.48/ 1.09±3.65 | 19 (27.27%) | 0.001164 ± 0.0016 |
| 9 | 22 | 1.0±0.44/ 0.95±3.09 | 7 (36.36%) | 0.001157 ± 0.0015 |
| 10 | 29 | 1.03±0.19/ 0.97±4.34 | 7 (34.48%) | 0.001173 ± 0.0023 |
| 11 | 32 | 0.97±0.18/ 0.97±5.48 | 24 (37.5%) | 0.001179 ± 0.003 |
| 12 | 15 | 1.0±0.65/ 0.93±2.58 | 7 (66.67%) | 0.001179 ± 0.0013 |
| 13 | 41 | 1.27±0.5/ 1.27±4.59 | 11 (29.27%) | 0.001148 ± 0.0019 |
| 14 | 19 | 0.95±0.23/ 0.95±4.13 | 24 (26.32%) | 0.001148 ± 0.0019 |
| 15 | 11 | 1.0±0.0/ 0.91±3.02 | 24 (81.82%) | 0.001129 ± 0.0016 |
| 16 | 48 | 1.1±0.47/ 1.19±6.58 | 5 (45.83%) | 0.001174 ± 0.003 |
| 17 | 32 | 1.0±0.0/ 0.97±5.48 | 19 (96.88%) | 0.001174 ± 0.003 |
| 18 | 19 | 0.95±0.23/ 0.95±4.13 | 9 (100.0%) | 0.001179 ± 0.0022 |
| 19 | 50 | 0.98±0.14/ 1.1±7.78 | 19 (50.0%) | 0.001177 ± 0.0039 |
| 20 | 8 | 0.88±0.35/ 0.88±2.47 | 18 (100.0%) | 0.001179 ± 0.0013 |
| 21 | 13 | 0.92±0.28/ 0.92±3.33 | 14 (38.46%) | 0.001179 ± 0.0018 |
| 22 | 29 | 0.97±0.19/ 1.0±5.39 | 6 (75.86%) | 0.001173 ± 0.0028 |
| 23 | 9 | 0.89±0.33/ 0.89±2.67 | 7 (66.67%) | 0.001179 ± 0.0014 |

TABLE III: Size (in nodes), average degree, major category, and major category coverage (in percent of nodes) in the identified communities in the HSOs network. The average degree and PageRank score of nodes within communities are quite evenly distributed across communities. A significant structural variation with respect to composition can be nevertheless observed. For instance, communities 1 and 18 show significantly (relative to the other communities) low and high concentration of nodes from a single category, respectively.

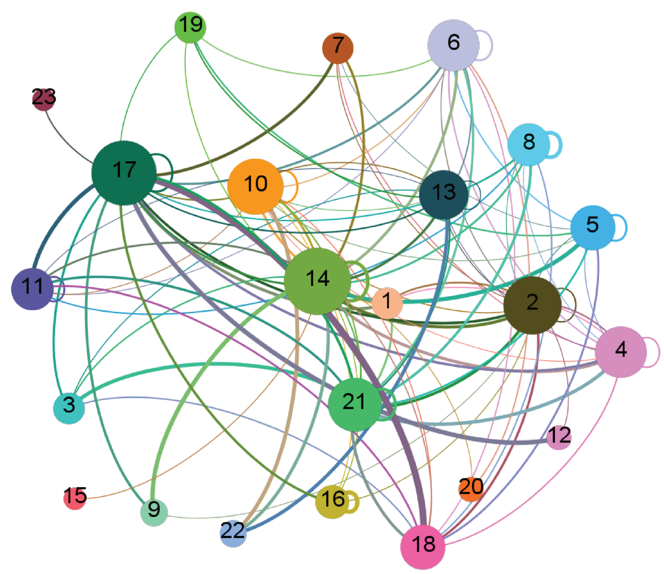


Fig. 5: Interactions between categories. Node sizes are in proportion to the size of each category, and edge thickness encodes the number of partnerships.

human services shows strong modularity with a score of 0.86 for a total of 23 communities, shown using different colors in Figure 6. The strong modularity is indicative of structurally independent communities, a precondition for multipolarity.

To test the multipolarity hypothesis, as opposed to a core–periphery structure, we consider (i) the average degree of nodes within a community, and (ii) the average PageRank score of each community, as indicators of structural importance. The results corroborate the multipolarity hypothesis (see Table III). Specifically, while the largest community has higher average degree and PageRank score than the other communities, communities appear to be equally important. We quantify this using the Gini coefficient [33]. The Gini coefficient is computed as the ratio between the area enclosed by the main diagonal (often called the “line of equality”) and the Lorenz curve, and the total triangular area under the line of equality. It varies between 0 and 1, with 0 indicating indicating perfect equality. The obtained value of 0.5159 for this network provides complimentary evidence in support of the multipolar structure of the HSOs network, with PageRank scores being almost evenly distributed across communities.

Finally, we compute for each community the fraction of nodes that belong to the most prevalent category as opposed to the rest of the nodes within the community. This provides

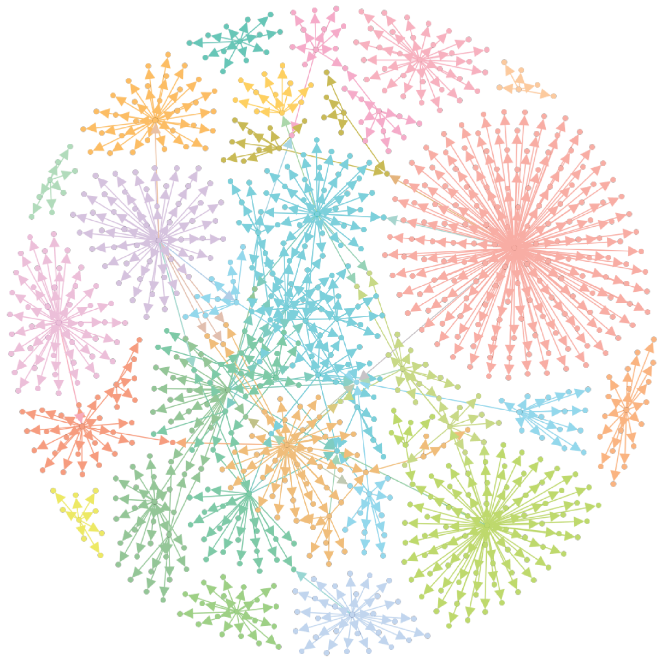

Fig. 6: The network of HSOs (better seen in color). The giant component encompasses 96.4% of the nodes. Colors denote communities nodes belong to (c.f. Section V).

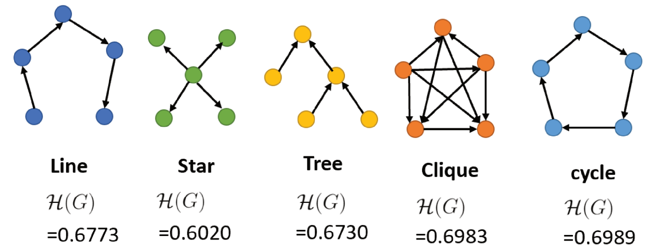


Fig. 7: Toy networks and their corresponding entropy. From left to right: chain, star, tree, clique, and circle.

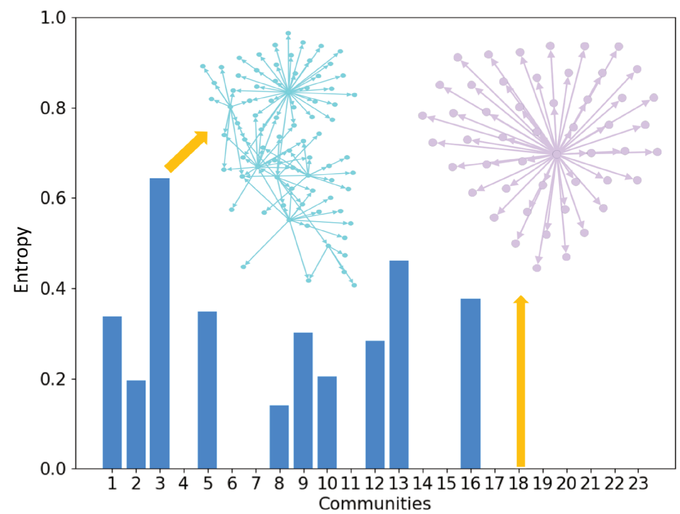


Fig. 8: Distribution of structural variation across the 23 communities in our network. The insets show the network structure of the community with the highest structural score and one of the communities with zero entropy, respectively.

a simplistic measure of community composition homogeneity. We find strong evidence that most of the communities are not homogeneous. Specifically, the majority of the communities comprise nodes from multiple categories. Nevertheless, extreme cases for which the fraction of nodes that belong to the most prevalent category is anomalously high or strikingly low, exist. For example, nodes from categories 8 (Crime & Legal-Related) and 19 (Science & Technology) span 100% of communities 18 and 20 respectively, whereas in community 1, even the most prevalent category (i.e., Education) accounts only for $\sim 21\%$ of the nodes. The existence of these extreme values in either direction does not lead to the rejection of the null hypothesis (i.e., that the fraction of nodes that belong to the most prevalent category is independent of community membership). The P–value of Fisher Exact Test is $9 \times 10^{-8}$, leading to the rejection of the null hypothesis.

Taken together, these results suggest that the network of human services is a fundamentally multipolar structure, where communities have roughly equal levels of structural importance, and are subject to similar (with few exceptions) level of composition heterogeneity. In future work, we plan to further explore this finding in relation to geographic boundaries to determine if ties between organizations are formed due to their close vicinity to one another, or irrespective of their location.

## VI. Structural Variation within Communities

In order to study structural variation across the 23 communities in $\mathcal{G}'$, we use a measure independent of the number of nodes in each community as follows. We begin by computing the Laplacian matrix $L = D - A$, where $A$ is the undirected adjacency matrix of $\mathcal{G}'$ and $D$ is the corresponding degree matrix. We then (i) compute $L$'s eigenvalues, which are real and non–negative since $L$ is symmetric, (ii) sort the eigenvalues in descending order $0 \leq \lambda_1 \leq \cdots \leq \lambda_{|V|}$, (iii) divide each eigenvalue by $\sum_i \lambda_i$ so that the sum of normalized eigenvalues becomes 1, and (iv) compute the entropy $\mathcal{H}(\mathcal{G}') = -\sum_{i=1} \lambda_i \log \lambda_i$. These steps are necessary (i.e., values have to be real, positive, and sum to one) for the entropy to be meaningful. For illustrative purposes, Figure 7 shows the entropy of toy networks. After calculating $\mathcal{H}$ across all 23 communities, we find it ranging from 0 to 0.6444 (see Figure 8). More than half of the communities have score 0, i.e., their topology resembles a star (e.g., inset for community 18 in Figure 8). No community approaches 1, indicating that circular or chain structures are not present. Nevertheless, communities with scores $\geq 0.2$, such as community 3, demonstrate a form of hierarchical, tree–like structure.

## VII. Conclusion

This study represents the first attempt to analyze the network of HSOs in Albany, the capital of the state of New York.

By collecting partnerships between HSOs from the Web, we constructed and analyzed their relationships network, and described some of its salient characteristics. Specifically, we analyzed the topological structure of this network to identify structurally important nodes and communities, and study the interactions between different categories of service–providing organizations. We discovered that unlike other social, technological and biological networks with a small–world structure, the network of HSOs is characterized by a multipolar structure with few super connectors, and strong relations between organizations that serve similar functions.

One possible criticism of our study is that it does not account for network evolution. In fact, given the notorious dynamism of the Web, there is no guarantee that a webpage available at any given point in time is going to be online indefinitely. In the context of human service organizations, lack of resources and funding may exacerbate the unavailability of information on the Web. In the face of such dynamism, one could argue that our snapshot of the network of HSOs from web data may not provide a good representation of the system. Our rationale for using the Web as our data source is that this may be the medium that people use to find and access relevant information while searching for service providers.

In future work, we plan to expand the set of organizations in our dataset beyond the area surounding Albany to the state of New York, beyond states in the U.S., and even beyond countries so as to get a more representative view of the networked human services. Second, we plan to extend our Web crawler so as to identify partnerships not only from textual data but also from images and PDF files. Third, we plan to examine and evaluate the stability of the network of HSOs over time. Finally, the paper only presented a subset of the possible analyses that can be performed using the collected data. In future work, we plan to examine multi–layer network representations to improve the insights into the complex network of human service organizations.

## References


[1] Y. Hasenfeld, *Human services as complex organizations*. Sage Publications, 2009.

[2] K. Snavely and M. B. Tracy, "Collaboration among rural nonprofit organizations," *Nonprofit Management and Leadership*, vol. 11, no. 2, pp. 145–165, 2000.

[3] K. G. Provan, M. A. Veazie, L. K. Staten, and N. I. Teufel-Shone, "The use of network analysis to strengthen community partnerships," *Public Administration Review*, vol. 65, no. 5, pp. 603–613, 2005.

[4] J. E. Austin, *The collaboration challenge: How nonprofits and businesses succeed through strategic alliances*. John Wiley & Sons, 2010, vol. 109.

[5] A. C. Bunger, "Administrative coordination in nonprofit human service delivery networks: The role of competition and trust," *Nonprofit and voluntary sector quarterly*, vol. 42, no. 6, pp. 1155–1175, 2013.

[6] P. Fredericksen and R. London, "Disconnect in the hollow state: The pivotal role of organizational capacity in community-based development organizations," *Public Administration Review*, vol. 60, no. 3, pp. 230–239, 2000.

[7] K. Kernaghan, "Moving towards the virtual state: Integrating services and service channels for citizen-centred delivery," *International Review of Administrative Sciences*, vol. 71, no. 1, pp. 119–131, 2005.

[8] M. Horvitz-Lennon, A. M. Kilbourne, and H. A. Pincus, "From silos to bridges: Meeting the general health care needs of adults with severe mental illnesses," *Health affairs*, vol. 25, no. 3, pp. 659–669, 2006.

[9] T. Corbett and J. Noyes, *Human services systems integration: A conceptual framework*. Institute for Research on Poverty, 2008.

[10] B. Chen, "Assessing interorganizational networks for public service delivery: A process-perceived effectiveness framework," *Public Performance & Management Review*, vol. 31, no. 3, pp. 348–363, 2008.

[11] K. Kelley, B. Clark, V. Brown, and J. Sitzia, "Good practice in the conduct and reporting of survey research," *International Journal for Quality in health care*, vol. 15, no. 3, pp. 261–266, 2003.

[12] W. Lee, "Sustainability of nonprofit human service organizations in a neighborhood context," *Nonprofit Management and Leadership*, vol. 28, no. 1, pp. 11–24, 2017.

[13] S. P. Borgatti and M. G. Everett, "Models of core/periphery structures," *Social networks*, vol. 21, no. 4, pp. 375–395, 2000.

[14] N. M. Tichy, M. L. Tushman, and C. Fombrun, "Social network analysis for organizations," *Academy of management review*, vol. 4, no. 4, pp. 507–519, 1979.

[15] N. Nohria, "Is a network perspective a useful way of studying organizations," *Hickman, Gill Robinson. Leading organizations: perspectives for a new era. California: Sage Publications*, pp. 287–301, 1998.

[16] D. Krackhardt, "Assessing the political landscape: Structure, cognition, and power in organizations," *Administrative science quarterly*, pp. 342–369, 1990.

[17] T. E. Stuart, "Network positions and propensities to collaborate: An investigation of strategic alliance formation in a high-technology industry," *Administrative science quarterly*, pp. 668–698, 1998.

[18] C. Guo and M. Acar, "Understanding collaboration among nonprofit organizations: Combining resource dependency, institutional, and network perspectives," *Nonprofit and voluntary sector quarterly*, vol. 34, no. 3, pp. 340–361, 2005.

[19] F. Wang, U. Srinivasan, S. Uddin, and S. Chawla, "Application of network analysis on healthcare," in *Proceedings of the 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. IEEE Press, 2014, pp. 596–603.

[20] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and analysis of online social networks," in *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*. ACM, 2007, pp. 29–42.

[21] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener, "Graph structure in the web," *Computer networks*, vol. 33, no. 1-6, pp. 309–320, 2000.

[22] E. O. Laumann, P. V. Marsden, and D. Prensky, "The boundary specification problem in network analysis," *Research methods in social network analysis*, vol. 61, p. 87, 1989.

[23] J. H. Levine, "The network of corporate interlocks in the United States: An overview," in *American Sociological Association Annual Meeting*, 1977.

[24] A. Clauset, C. R. Shalizi, and M. E. Newman, "Power-law distributions in empirical data," *SIAM review*, vol. 51, no. 4, pp. 661–703, 2009.

[25] S. H. Strogatz, "Exploring complex networks," *nature*, vol. 410, no. 6825, p. 268, 2001.

[26] L. A. Adamic, R. M. Lukose, A. R. Puniyani, and B. A. Huberman, "Search in power-law networks," *Physical review E*, vol. 64, no. 4, p. 046135, 2001.

[27] D. S. Callaway, M. E. Newman, S. H. Strogatz, and D. J. Watts, "Network robustness and fragility: Percolation on random graphs," *Physical review letters*, vol. 85, no. 25, p. 5468, 2000.

[28] A. Bavelas, "Communication patterns in task-oriented groups," *The Journal of the Acoustical Society of America*, vol. 22, no. 6, pp. 725–730, 1950.

[29] P. Bonacich, "Power and centrality: A family of measures," *American journal of sociology*, vol. 92, no. 5, pp. 1170–1182, 1987.

[30] M. Girvan and M. E. Newman, "Community structure in social and biological networks," *Proceedings of the national academy of sciences*, vol. 99, no. 12, pp. 7821–7826, 2002.

[31] S. Zhou and R. J. Mondragón, "The rich-club phenomenon in the Internet topology," *IEEE Communications Letters*, vol. 8, no. 3, pp. 180–182, 2004.

[32] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of statistical mechanics: theory and experiment*, vol. 2008, no. 10, p. P10008, 2008.

[33] A. B. Atkinson *et al.*, "On the measurement of inequality," *Journal of economic theory*, vol. 2, no. 3, pp. 244–263, 1970.