

Centrality in Collaboration: A Novel Algorithm for Social Partitioning Gradients in Community Detection for Multiple Oncology Clinical Trial Enrollments

Benjamin Smith¹, Tyler Pittman², and Wei Xu^{1,2}

¹University of Toronto, Toronto, Canada

²University Health Network, Toronto, Canada

Abstract

Patients at a comprehensive cancer center who do not achieve cure or remission following standard treatments often become candidates for clinical trials. Patients who participate in a clinical trial may be suitable for other studies. A key factor influencing patient enrollment in subsequent clinical trials is the structured collaboration between oncologists and most responsible physicians. Possible identification of these collaboration networks can be achieved through the analysis of patient movements between clinical trial intervention types with social network analysis and community detection algorithms. In the detection of oncologist working groups, the present study evaluates three community detection algorithms: Girvan-Newman, Louvain and an algorithm developed by the author. Girvan-Newman identifies each intervention as their own community, while Louvain groups interventions in a manner that is difficult to interpret. In contrast, the author's algorithm groups interventions in a way that is both intuitive and informative, with a gradient evident in social partitioning that is particularly useful for epidemiological research. This lays the groundwork for future subgroup analysis of clustered interventions.

1 Introduction

When cancer patients complete standard treatments at Princess Margaret Cancer Centre and have not responded with being cured or in remission, they become candidates for clinical trials. These clinical trials are regulated studies registered by Health Canada¹ as opposed to quality assurance studies². Patients who qualify may have been screen failures for other trials, have experienced progressive disease, or are receiving maintenance therapy and have been referred to a clinical trial by their oncologist or most responsible physician. Ground truth shows that collaboration networks between oncologists is a primary factor for further engagement in subsequent clinical trials by patients after completion of the given clinical trial that they are enrolled in. A possible approach to understanding the structure of these collaboration networks is through use of social network analysis (SNA) and community detection algorithms.

Social network analysis examines individual entities and their relationships among them. The data is represented as a “graph” where individual entities are referred to as “nodes” and their relationships between them as “edges”, which may be directional if specified (see Figure 1). A primary area of study in SNA is the analysis of interconnectivity of nodes, called “communities” and identification of clusters through the use of algorithms called “community detection algorithms”. Rostami et al¹ (2023) note that there is no specific model which describes exactly what a “community” is. Generally, community detection algorithms employ specific optimization strategies to partition a large-scale complex network into a set of disjoint and compact subgroups, often (but not always) without prior knowledge regarding the number of subgroups and their sizes. Rostami et al further note that it is commonly acknowledged that there is no unique community detection algorithm that can accommodate all kinds of graphs, because of the inherent variability in network structures and their respective objective(s).

Application of community detection algorithms with oncology clinical trial data has been preformed in the past. Georgiev et al² (2011) applied the Girvan-Newman³ (2002) algorithm and noted a lack of cohesion among researchers who studied treatments for multiple myeloma. Haq and Wang⁴ (2016) applied the Louvain algorithm (by Blondel et al⁵ (2008)) to identify communities of cancer patients with significantly different survival curves. The present study applies SNA, and compares multiple community detection algorithms to identify collaboration networks between oncologists through the interventions studied in clinical trials via enrollment data of patients in multiple, nonconcurrent clinical trials at Princess Margaret Cancer Centre (PM) in Toronto, Canada. Inspired by work from Gorgiev et al (2011), Haq and Wang (2016), Ostovari and Yu⁶ (2019) and Bissoyi and Patra⁷ (2020) this research considers the Girvan-

¹For more information, see <https://www.canada.ca/en/health-canada/services/clinical-trials.html>

²Quality assurance studies in the context of medical studies are studies which look at drugs which are already approved for use, but the goals are focused on other aspects of care such as drug delivery or quality of care.

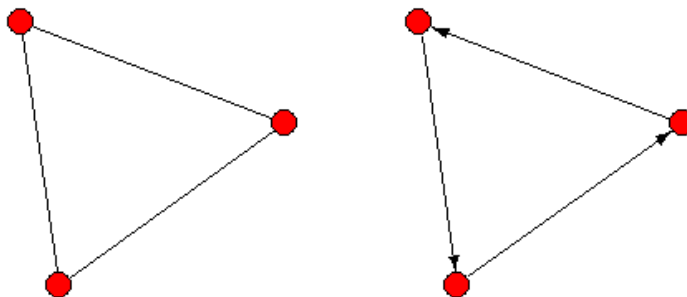


Figure 1: Two simple graphs with directed and undirected edges. Direction is noted by arrowheads at the end of the edges.

Newman and Louvain algorithms and compares them to an author-developed algorithm, referred to as “Smith-Pittman”³, to identify collaboration networks between clinical trials classified by intervention.

2 Materials and Methods

2.1 The Data

The data is a result of multiple data source integration, which was undertaken from the PM Cancer Registry and Clinical Research Record. The data is anonymized, and spans patient enrollments in oncology clinical trials between January 1, 2016 and December 31, 2018. In this time period, there were 2970 patients enrolled in 515 clinical trials involving 41 principal investigators. For the identification of collaboration networks between oncologists, the analytic sample only consists of patients who were enrolled in more than one clinical trial within the time period studied. The resulting analytic sample consists of 389 patients enrolled in 288 clinical trials. Among these clinical trials, some interventions can be classified into broader categories of targeted therapies, or immunotherapy. This has been identified in the data with “T:” and “I:” prefixes respectively. The clinical trials were classified by intervention type, presenting as 16 distinct intervention types among 470 patient enrollments. With this classification, the patient referral graph is constructed (see Figure 2).

The analysis is preformed with the R programming language, and makes use of an extensive array of libraries and dependencies. The primary libraries that were utilized include `igraph`, `tidyverse`, and `tidygraph`. For the complete script, please refer to the Appendix - Program Syntax.

³Named after the author and his co-supervisor, Tyler Pittman.

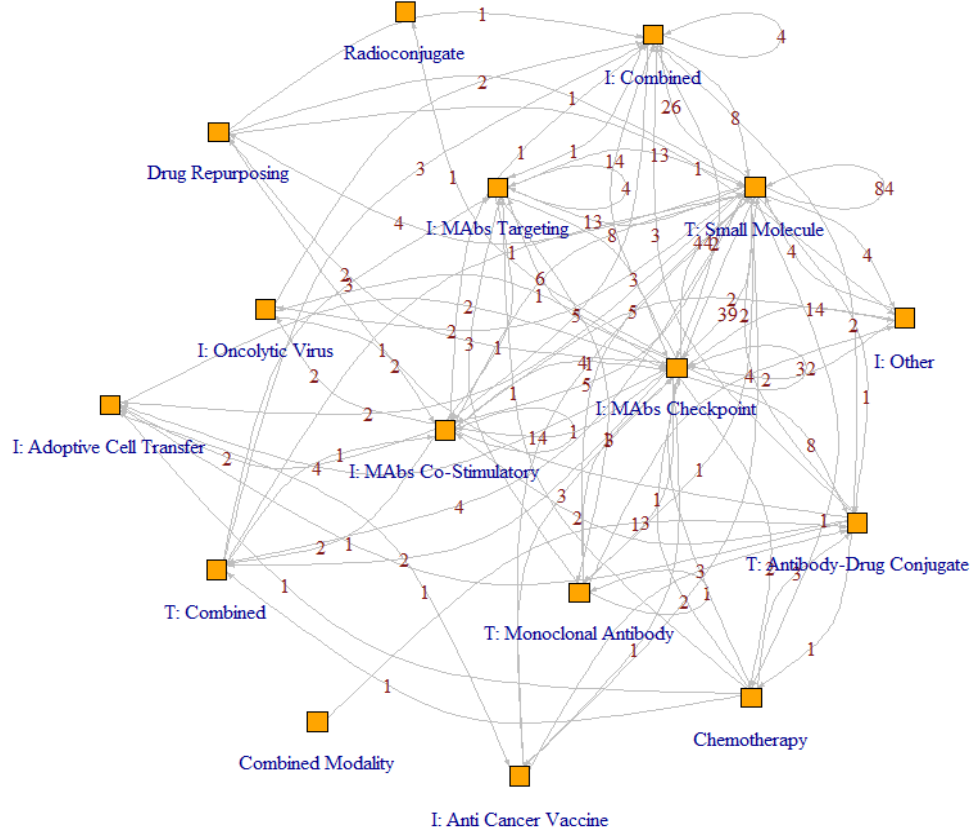


Figure 2: Patient movement between clinical trials classified by intervention type at PM. Nodes indicate the treatment type, and labeled edges indicate the movement (subsequent enrollment) of patients between clinical trials in a given intervention of the same type (self loop), or differing. Among the clinical trials, some interventions can be classified into broader categories consisting of targeted therapies or immunotherapy. This has been identified in the data with “T:” and “I:” prefixes respectively.

2.2 Methods

The goal of applying community detection algorithms with this data is to identify oncologist working groups among treatment interventions, based on the movement (incoming and outgoing referrals) of patients between the intervention types. These movements in the network are understood through measures that are considered by the community detection algorithms' optimization strategies. While the Girvan-Newman, Louvain and Smith-Pittman algorithms differ in their approaches to the identification of collaboration networks, their identification strategies are based on the maximization of modularity, Q - a measure that scores the degree of segregation within a network through tightly connected communities or clusters (See Newman⁸ (2006)).

The mathematical representation of modularity is defined in the `igraph` R package⁹ (2006) as:

$$Q = \frac{1}{m} \sum_{i,j} \left(A_{ij} - \frac{k_i^{\text{out}} k_j^{\text{in}}}{m} \right) \delta(c_i, c_j)$$

Where m is the number of edges (patient movements), A_{ij} is the number of connections shared by nodes i and j (movements between interventions i and j), k_i^{out} and k_j^{in} are the number of edges coming out from node i and going into node j (patient movements from intervention i and j) and $\delta(c_i, c_j)$ is an indicator variable identifying if nodes i and j are connected - either directly or through another node (if there is a patient movement between interventions i and j either directly or through some other intervention). For directed graphs, k_i^{out} and k_j^{in} are simply the number of connected edges possessed by nodes i and j , respectfully. For a more comprehensive overview modularity and other measures in social network analysis, see Newman (2006), Wasserman and Faust¹⁰ (1994) and Latora et al¹¹ (2017).

2.2.1 Girvan-Newman

The Girvan-Newman algorithm is based on the evaluation of edges in a social network through edge-betweenness centrality. Edge-betweenness centrality is defined by Girvan and Newman (2002) as the number of shortest paths that go through an edge in a graph, divided by the total number of shortest paths between nodes i and j . Each edge in a graph has its own edge-betweenness centrality value. The `igraph` (2006) documentation defines edge-betweenness centrality for an edge e in a social network in mathematical terms as:

$$\sum_{i \neq j} g_{iej} / g_{ij}$$

Where g_{ij} is the number of shortest paths between nodes i and j (patient movements between interventions i and j , either directly or through some other intervention(s)), and g_{iej} is the number of shortest paths which pass through

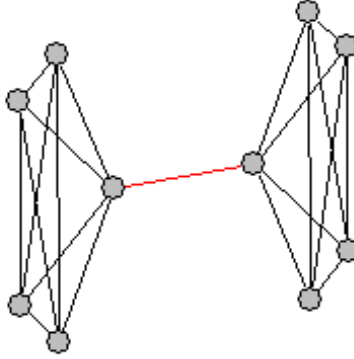


Figure 3: A simple network demonstrating an edge with a high edge-betweenness centrality, highlighted in red. The network consists of two densely connected clusters, with the red edge serving as the sole connection between them. This edge is crucial for communication between the two clusters, as most of the shortest paths that connect nodes from opposite clusters pass through it.

edge e . Figure 3 provides an illustration of a simple network, showing the edge with the highest edge-betweenness centrality highlighted in red.

Edge betweenness can be calculated for directed and undirected edges. As a result, the Girvan-Newman algorithm can be applied to directed or undirected graphs without any transformations. The steps of the Girvan-Newman algorithm are as follows:

1. Calculate edge-betweenness centrality for all edges in the network.
2. Find the edge with the highest edge-betweenness centrality, and remove it from the network.
3. Recalculate edge-betweenness centrality for all remaining edges.
4. Repeat from step 2.

Girvan-Newman can be utilized when the community structure is known, and will classify nodes into a predetermined number of communities based on the hierarchy produced by the algorithm (see Girvan and Newman 2002). When the community structure is not known, modularity is evaluated after each iteration of the algorithm. The grouping of nodes into distinct communities is selected via modularity maximization.

2.2.2 Louvain

The Louvain algorithm (by Blondel et al 2008) operates in two distinct phases. (i) In the first phase, each node in the network is considered as their own community, resulting in the initial partition with as many communities as there

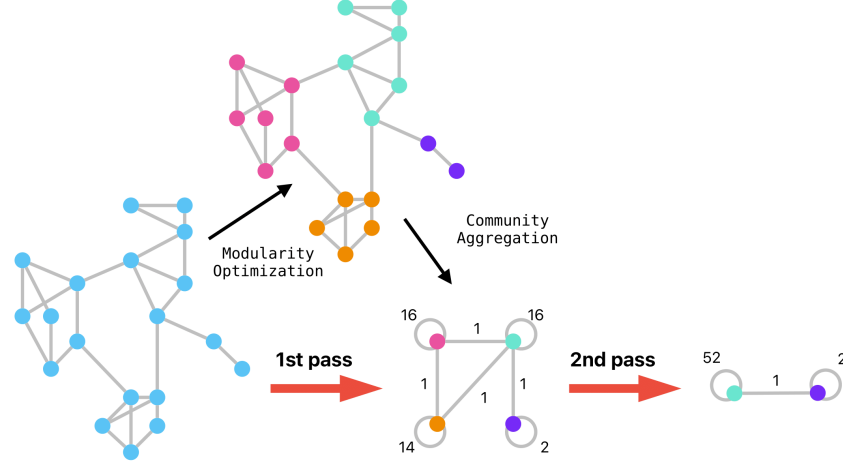


Figure 4: Reproduced illustration of the Louvain algorithm (originally designed by Blondel et al (2008)).

are nodes. The algorithm then assesses the potential modularity gain for each node i if it were to leave its current community and join the community of node j . After evaluating the potential modularity gain across all communities, node i is reassigned to the community of node j , where the modularity increase is maximized. The process is iteratively and sequentially applied for all nodes, until no further improvement can be achieved. This first phase stops when a local maximum of modularity is reached, meaning that no subsequent node move can enhance modularity. (ii) The second phase involves constructing a new network as represented by the communities identified in the first phase. Links between nodes of the same community are viewed as “self-loops” for the community in the new network. Once this second phase is complete, the first phase of the algorithm can be reapplied. The combination of these two phases is referred to as a “pass”. The algorithm terminates when there is no other local maxima in modularity to be achieved in subsequent passes.

A key limitation of the Louvain algorithm is that it is generally programmed to work only with undirected graphs⁴. In order to apply the Louvain algorithm to a directed graph, it must first be converted to an undirected graph. Figure 4 is a reproduction of Blondel et al’s (2008) illustration of the algorithm.

⁴Work on extending the Louvain algorithm to accommodate directed graphs has been an outstanding issue in the igraph developer community since 2015 (See: <https://github.com/igraph/igraph/issues/890>). However, Dugué and Perez¹² (2022) have done some work on this.

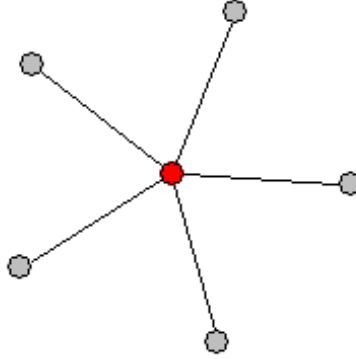


Figure 5: A simple network highlighting node degree. The center node (colored red) possesses the highest number of connections and as a result possesses the highest degree and degree centrality index.

2.2.3 Smith-Pittman

The "Smith-Pittman" algorithm is a modification of the Girvan-Newman algorithm, where degree centrality is considered. Degree centrality of a node is simply defined as the number of connections a node has within a given network (see figure 5). The algorithm proceeds through the following steps:

1. Calculate the degree centrality for each node, and the edge-betweenness centrality of all edges in the network.
2. Identify the subgraph associated with the node that has the highest degree centrality.
3. Remove the edge possessing the highest calculated edge-betweenness centrality.
4. Recalculate the degree centrality for all nodes, and the edge-betweenness centrality for the remaining edges in the network.
5. Repeat from step 2.

Figure 6 provides a visual representation of this algorithm. Like Girvan-Newman, the Smith-Pittman algorithm can be applied to both directed and undirected graphs. Conceptually, the algorithm can be specified to terminate once a predetermined number of communities have been identified. However, its primary design is for use in an unsupervised setting, where clusters are identified through the maximization of modularity as evaluated after each iteration of the algorithm.

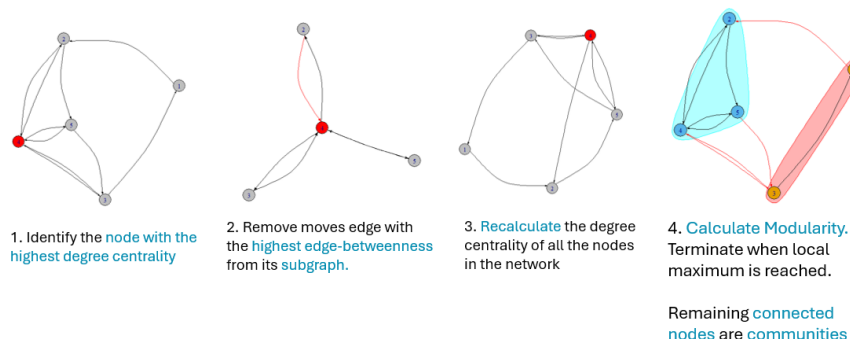


Figure 6: Illustration of the Smith-Pittman algorithm. Highlighted convex hulls denote the identification of distinct communities.

3 Results

Figures 7-9 show the communities identified by the algorithms, through convex hulls highlighting the grouped interventions. Tables 1-3 show the grouping of interventions into communities, and the breakdown by frequency of incoming and outgoing patient referrals for each treatment intervention studied. Figure 7 demonstrates that the Girvan-Newman algorithm identified each intervention as a separate community ($Q = 0.044$). This result is particularly uninformative, as it is equivalent to not applying any community detection method to identify oncologist collaboration networks between the interventions. Figure 8 shows that the Louvain algorithm groups interventions into four distinct working groups, achieving the highest modularity score ($Q = 0.177$). However, the underlying rationale and meaning behind these groupings remains unclear, beyond the objective to cluster interventions as to maximize modularity.

Figure 9 shows that the Smith-Pittman algorithm ($Q = 0.08$) identified eight communities. Six of these communities consist of individual interventions - namely T: Small Molecule, I:MAbs⁵ Checkpoint, I:Combined, I:MAbs Targeting, Combined Modality and Radioconjugate - while the remaining two communities encompass multiple interventions. The interpretation of the communities identified by the Smith-Pittman algorithm can be facilitated by the degree of connectivity among the interventions within these communities. Communities comprised of individual interventions either have the highest or a substantial number of patient referrals, whether incoming from or outgoing to other interventions, or they have the least. Figure 10 illustrates the distribution of interventions by patient referrals, ordered from smallest to largest, and highlights the thresholds beyond which single intervention communities are positioned. The interpretation of the communities identified by the Smith-Pittman algorithm suggests the existence of both highly connected, and less connected interventions, as well as broader groups corresponding to typical intervention

⁵Short for Monoclonal Antibodies.

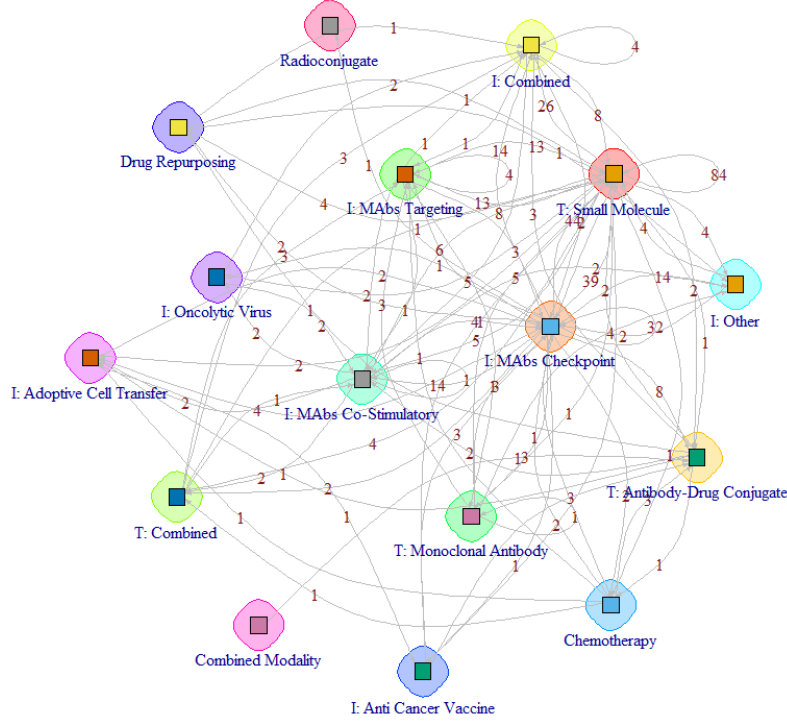


Figure 7: Detected communities via Girvan-Newman with modularity maximization. Sixteen distinct communities.

types - there is a gradient that is evident in social partitioning. This interpretation offers an intuitive understanding related to the formation of collaboration networks being a function of intervention “popularity” - i.e. patient referrals outgoing and incoming to and from other interventions.

4 Discussion

Where the Girvan-Newman algorithm failed to identify communities, the Louvain and Smith-Pittman algorithms succeeded. A possible explanation for this discrepancy lies in the nature of the data analyzed, which includes patient referrals to clinical trials that investigate the same intervention types as the clinical trials patients were previously enrolled in. In graph theory, such referrals are represented as “self loops” and introduce complexity in the network. The Girvan-Newman algorithm - whose original design was not for complex networks - struggles in such contexts, leading to its failure to group multiple interventions into communities based on modularity maximization.

The Louvain algorithm successfully detected communities. However, the re-

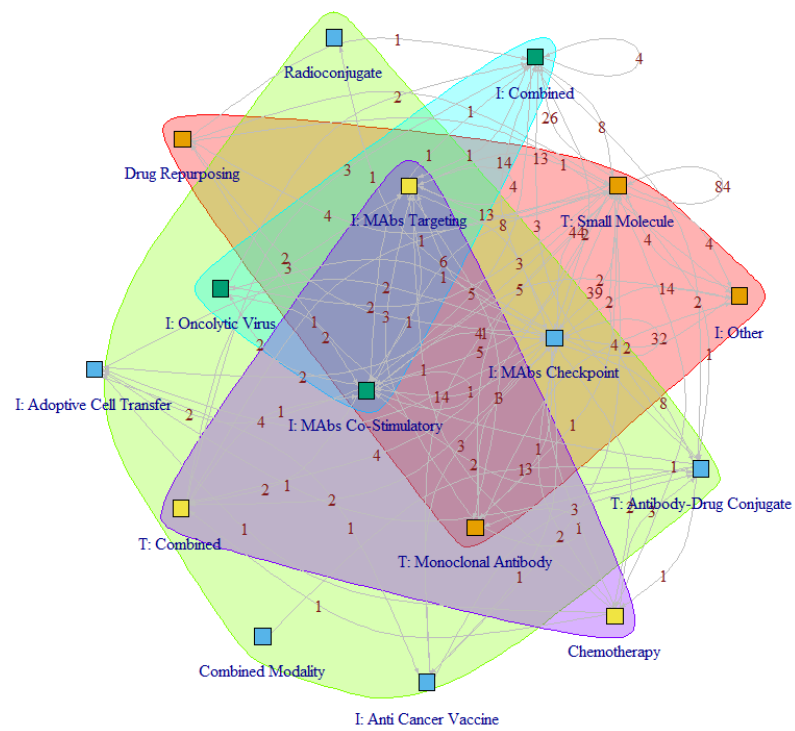


Figure 8: Detected communities via Louvain algorithm with modularity maximization. Four distinct communities.

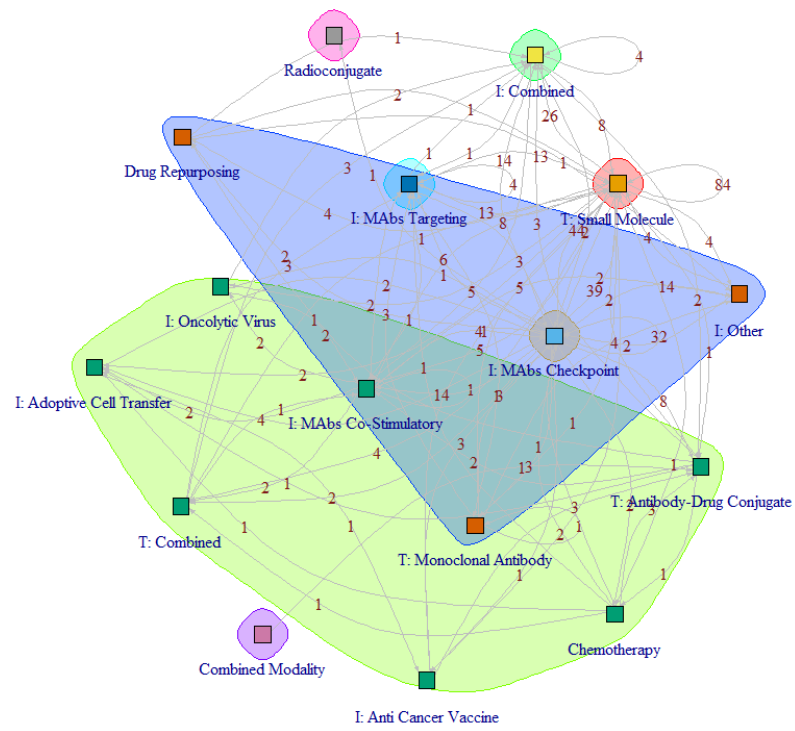


Figure 9: Detected communities via Smith-Pittman algorithm with modularity maximization. Eight distinct communities.

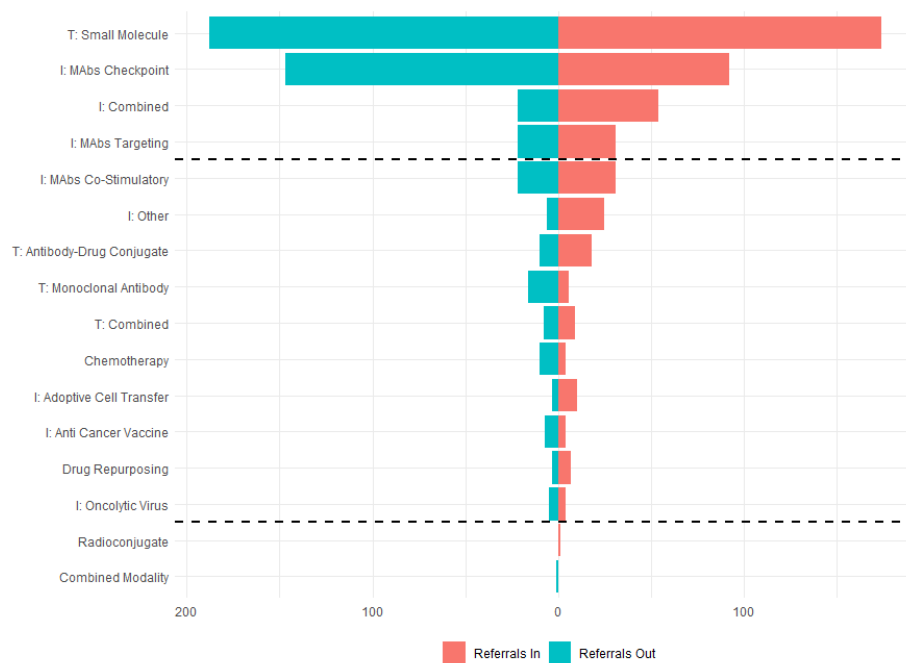


Figure 10: Referral distribution among interventions. Interventions outside the boundaries (T: Small Molecule, I:MAb's Checkpoint, I: Combined, I:Mab's Targeting, Radioconjugate and Combined Modality) are each identified as individual communities, while interventions within them are identified as belonging to communities consisting of multiple interventions.

Intervention	Referrals In	Referrals Out	Total
Chemotherapy	4	10	14
Combined Modality	0	1	1
Drug Repurposing	7	3	10
I: Adoptive Cell Transfer	10	3	13
I: Anti Cancer Vaccine	4	7	11
I: Combined	54	22	76
I: MAbs Checkpoint	92	147	239
I: MAbs Co-Stimulatory	31	22	53
I: MAbs Targeting	31	22	53
I: Oncolytic Virus	4	5	9
I: Other	25	6	31
Radioconjugate	1	0	1
T: Antibody-Drug Conjugate	18	10	28
T: Combined	9	8	17
T: Monoclonal Antibody	6	16	22
T: Small Molecule	174	188	362

Table 1: Girvan-Newman communities identified. Each intervention is their own community.

sulting groups were difficult to interpret. This difficulty arises because the Louvain algorithm bases its community selection purely on modularity maximization, and does not consider the direction of patient movements the underlying structural or functional significance of particular interventions in the context of the network. The primary advantage of the Louvain algorithm is its efficiency in performing community detection on large networks. It has been widely used in applications such as the Twitter Social Network (Pujol et al. 2009) which consisted of 2.4 million nodes and 38 million links, and mobile phone network data (Greene et al. 2010) with 4 million nodes, 100 million links. These networks are orders of magnitude larger than the patient referral network analyzed in this study, highlighting the scalability of the Louvain algorithm. However, utility of such a algorithm is limited in smaller, more specialized networks where the interpretability and justification of communities identified is important.

In contrast, the Smith-Pittman algorithm directly addresses connectivity of interventions studied in the clinical trials, by incorporating degree centrality and edge-betweenness centrality. This approach allows for the identification of communities with a more ordered structure, distinguishing between highly connected and minor interventions as they reflect the relational dynamics in the network. The results from the Smith-Pittman algorithm are promising, however, the results from this analysis alone is insufficient to establish generalizability of the algorithm. To fully assess its usefulness, a formal simulation study with further application of the Smith-Pittman algorithm in diverse settings is necessary. Additionally, the practical value of identified communities will become evident

Intervention	Referrals In	Referrals Out	Total
Community: 1			
Drug Repurposing	7	3	10
I: Other	25	6	31
T: Monoclonal Antibody	6	16	22
T: Small Molecule	174	188	362
Community: 2			
Combined Modality	0	1	1
I: Adoptive Cell Transfer	10	3	13
I: Anti Cancer Vaccine	4	7	11
I: MAbs Checkpoint	92	147	239
Radioconjugate	1	0	1
T: Antibody-Drug Conjugate	18	10	28
Community: 3			
I: Combined	54	22	76
I: MAbs Co-Stimulatory	31	22	53
I: Oncolytic Virus	4	5	9
Community: 4			
Chemotherapy	4	10	14
I: MAbs Targeting	31	22	53
T: Combined	9	8	17

Table 2: Louvain communities identified and grouped interventions

Intervention	Referrals In	Referrals Out	Total
Community: 1			
T: Small Molecule	174	188	362
Community: 2			
I: MAbs Checkpoint	92	147	239
Community: 3			
Chemotherapy	4	10	14
I: Adoptive Cell Transfer	10	3	13
I: Anti Cancer Vaccine	4	7	11
I: MAbs Co-Stimulatory	31	22	53
I: Oncolytic Virus	4	5	9
T: Antibody-Drug Conjugate	18	10	28
T: Combined	9	8	17
Community: 4			
I: Combined	54	22	76
Community: 5			
I: MAbs Targeting	31	22	53
Community: 6			
Drug Repurposing	7	3	10
I: Other	25	6	31
T: Monoclonal Antibody	6	16	22
Community: 7			
Combined Modality	0	1	1
Community: 8			
Radioconjugate	1	0	1

Table 3: Smith-Pittman communities and identified and grouped interventions

when they are applied as grouping variables in downstream analysis, such as outcome prediction or intervention effectiveness studies.

Further research should focus on subgroup analysis, and exploring extensions back to traditional statistical methods, such as regression and survival analysis. This research can further validate the utility of the identified communities, and the use of SNA and community detection algorithms in clinical research settings. The results of the Smith-Pittman algorithm lay the groundwork for these efforts, and potentially offer a robust tool for community detection in social and complex networks. Further work with the identified communities should involve assessment of the impact of community structure on patient outcomes, and identify if there are any structural inequities present in clinical trial enrollments. This line of research can lead to the identification of collaboration networks that improve patient care in clinical settings.

5 References

1. Rostami, M., Oussalah, M., Berahmand, K. & Farrahi, V. Community Detection Algorithms in Healthcare Applications: A Systematic Review. *IEEE Access* 11, 30247–30272 (2023).
2. Georgiev, H., Tsalatsanis, A., Kumar, A. & Djulbegovic, B. Social Network Analysis (SNA) of Research Programs In Multiple Myeloma (MM). *Blood* 118, 3144–3144 (2011).
3. Girvan, M. & Newman, M. E. J. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences* 99, 7821–7826 (2002).
4. Haq, N. & Wang, Z. J. Community detection from genomic datasets across human cancers. in 2016 IEEE Global Conference on Signal and Information Processing (GlobalSIP) 1147–1150 (IEEE, 2016). doi:10.1109/GlobalSIP.2016.7906021.
5. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* 2008, P10008 (2008).
6. Ostovari, M. & Yu, D. Impact of care provider network characteristics on patient outcomes: Usage of social network analysis and a multi-scale community detection. *PLoS One* 14, e0222016 (2019).
7. Bissoyi, S. & Patra, M. R. Community Detection in a Patient-Centric Social Network. in 171–182 (2021). doi:10.1007/978-981-15-7394-1_17.
8. Newman, M. E. J. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences* 103, 8577–8582 (2006).
9. Csardi, G. & Nepusz, T. The igraph software package for complex network research. *InterJournal, Complex Systems* 1695 (2006).
10. Wasserman, S. & Faust, K. *Social Network Analysis*. (Cambridge University Press, 1994). doi:10.1017/CBO9780511815478.
11. Latora, V., Nicosia, V. & Russo, G. *Complex Networks*. (Cambridge University Press, 2017). doi:10.1017/9781316216002.
12. Dugué, N. & Perez, A. Direction matters in complex networks: A theoretical and applied study for greedy modularity optimization. *Physica A: Statistical Mechanics and its Applications* 603, 127798 (2022).

6 Appendix - Program Syntax

```
1 library(tidygraph)
2 library(igraph)
3 library(ig.degree.betweenness) # Author developed
  methodology, pending public release
4 library(plyr) # for join_all
5 library(gt) # for tables
6 # Load R Data
7 real_df <- readRDS("path/to/data.rds")
8
9 real_df$New_Intervention_Name[real_df$New_Intervention_Name
  %in% c("Immunotherapy-
  MAbs-immunomodulatory-Checkpoint")] <- "I: MAbs
  Checkpoint";
10 real_df$New_Intervention_Name[real_df$New_Intervention_Name
  %in% c("Tageted therapy- antibody-drug conjugate")] <-
  "T: Antibody-Drug Conjugate";
11 real_df$New_Intervention_Name[real_df$New_Intervention_Name
  %in% c("Immunotherapy-
  MAbs-immunomodulatory-Co-Stimulatory")] <- "I: MAbs
  Co-Stimulatory";
12 real_df$New_Intervention_Name[real_df$New_Intervention_Name
  %in% c("Immunotherapy- Immuno + other investigational
  agent")] <- "I: Combined";
13 real_df$New_Intervention_Name[real_df$New_Intervention_Name
  %in% c("Targeted therapy - combined (small molecule +
  monoclonal antibody)")] <- "T: Combined";
14 real_df$New_Intervention_Name[real_df$New_Intervention_Name
  %in% c("Immunotherapy- MAbs- Tumour-targeting (includes
  immunoconjugates, naked MAbs)")] <- "I: MAbs Targeting";
15 real_df$New_Intervention_Name[real_df$New_Intervention_Name
  %in% c("Targeted therapy - small molecule")] <- "T:
  Small Molecule";
16 real_df$New_Intervention_Name[real_df$New_Intervention_Name
  %in% c("Immunotherapy- Other")] <- "I: Other";
17 real_df$New_Intervention_Name[real_df$New_Intervention_Name
  %in% c("Targeted therapy - monoclonal antibody")] <- "T:
  Monoclonal Antibody";
18 real_df$New_Intervention_Name[real_df$New_Intervention_Name
  %in% c("Immunotherapy- Adoptive Cell Transfer (e.g.
  TILS)")] <- "I: Adoptive Cell Transfer";
19 real_df$New_Intervention_Name[real_df$New_Intervention_Name
  %in% c("Immunotherapy- combined types")] <- "I:
  Combined";
20 real_df$New_Intervention_Name[real_df$New_Intervention_Name
  %in% c("Other - drug repurposing")] <- "Drug
  Repurposing";
```

```

21 real_df$New_Intervention_Name[real_df$New_Intervention_Name
    %in% c("Immunotherapy- Cytokines (eg. INFa, IL,
    Hematopoietic growth factors)")] <- "I: MAbs
    Co-Stimulatory";
22 real_df$New_Intervention_Name[real_df$New_Intervention_Name
    %in% c("Multiple- Biomarker Targeted")] <- "T: Combined";
23 real_df$New_Intervention_Name[real_df$New_Intervention_Name
    %in% c("Immunotherapy- Anti Cancer Vaccine- Peptide
    based vaccine")] <- "I: Anti Cancer Vaccine";
24 real_df$New_Intervention_Name[real_df$New_Intervention_Name
    %in% c("Chemotherapy")] <- "Chemotherapy";
25 real_df$New_Intervention_Name[real_df$New_Intervention_Name
    %in% c("Immunotherapy- Oncolytic Virus")] <- "I:
    Oncolytic Virus";
26 real_df$New_Intervention_Name[real_df$New_Intervention_Name
    %in% c("Combined modality (e.g chemoradiation,
    EBRT+Brachy)")] <- "Combined Modality";
27 real_df$New_Intervention_Name[real_df$New_Intervention_Name
    %in% c("Immunotherapy- Anti Cancer Vaccine- Gene Therapy
    (e.g DNA/RNA vaccines)")] <- "I: Anti Cancer Vaccine";
28 real_df$New_Intervention_Name[real_df$New_Intervention_Name
    %in% c("Other - radioconjugate")] <- "Radioconjugate";
29 real_df$New_Intervention_Name[real_df$New_Intervention_Name
    %in% c("Homonal Treatment")] <- "Drug Repurposing";
30
31
32 intervention_graph_real_directed<- real_df |>
33   dplyr::group_by(Subject_ID,Study_ID) |>
34   dplyr::filter(dplyr::n() > 1) |>
35   dplyr::distinct(pick(Subject_ID,Study_ID),.keep_all =
    TRUE) |>
36   dplyr::ungroup() |>
37   dplyr::filter(Subject_ID %in%
    names(table(Subject_ID))[table(Subject_ID) > 1]) |>
38   dplyr::group_by(Subject_ID) |>
39   dplyr::group_split() |>
40   lapply(function(x) x |>
41     dplyr::mutate(x, index = 1:nrow(x),
42                   direction = ifelse(index%%2 == 1,
43                                       "from","to"))) |>
43   do.call(what = rbind) |>
44   dplyr::select(Subject_ID, Study_ID, direction,
    New_Intervention_Name) |>
45   tidyr::pivot_wider(
46     id_cols = c(Subject_ID),
47     names_from = direction,
48     values_from = c(New_Intervention_Name, Study_ID)) |>
49   dplyr::rename(from = New_Intervention_Name_from,
50                 to = New_Intervention_Name_to,
51                 Study_ID = Study_ID_from) |>

```

```

52   tidy::unnest(from) |>
53   tidy::unnest(to) |>
54   tidy::unnest(Study_ID) |>
55   tidy::unnest(Study_ID_to) |>
56   dplyr::mutate(from = str_wrap(from, width = 30),
57                 to = str_wrap(to, width = 30)) |>
58   #dplyr::group_by(from, to) |>
59   #dplyr::count(name="Num_Patients") |>
60   tidygraph::as_tbl_graph(directed = TRUE) |>
61   igraph::as.igraph()
62
63
64   intervention_graph_real_undirected<- real_df |>
65   dplyr::group_by(Subject_ID) |>
66   dplyr::filter(dplyr::n() > 1) |>
67   dplyr::distinct(pick(Subject_ID, Study_ID), .keep_all =
68     TRUE) |>
69   dplyr::ungroup() |>
70   dplyr::filter(Subject_ID %in%
71     names(table(Subject_ID))[table(Subject_ID) > 1]) |>
72   dplyr::group_by(Subject_ID) |>
73   dplyr::group_split() |>
74   lapply(function(x) x |>
75     dplyr::mutate(x, index = 1:nrow(x),
76                   direction = ifelse(index%%2 == 1,
77                                     "from", "to"))) |>
75   do.call(what = rbind) |>
76   dplyr::select(Subject_ID, Study_ID, direction,
77     New_Intervention_Name) |>
78   tidy::pivot_wider(
79     id_cols = c(Subject_ID),
80     names_from = direction,
81     values_from = c(New_Intervention_Name, Study_ID)) |>
82   dplyr::rename(from = New_Intervention_Name_from,
83     to = New_Intervention_Name_to,
84     Study_ID = Study_ID_from) |>
85   tidy::unnest(from) |>
86   tidy::unnest(to) |>
87   tidy::unnest(Study_ID) |>
88   tidy::unnest(Study_ID_to) |>
89   dplyr::mutate(from = str_wrap(from, width = 30),
90     to = str_wrap(to, width = 30)) |>
91   tidygraph::as_tbl_graph(directed = FALSE) |>
92   igraph::as.igraph()
93
94   own_subj_mult_studies_check <- real_df |>
95   dplyr::distinct(Subject_ID, Study_ID) |>
96   dplyr::group_by(Subject_ID) |>
97   dplyr::count(name="N_Studies") |>

```

```

98     dplyr::filter(N_Studies > 1)
99 #389 participants enrolled in more than 1 study in 470
    instances;
100
101
102
103 ### Limit analysis to participants who enrolled in more
    than 1 clinical trial;
104
105
106 own <- real_df |>
107   # Adding this line because Tyler has it as well.
108   dplyr::mutate(eligible = "eligible") |>
109   dplyr::filter(Subject_ID %in%
110     unique(own_subj_mult_studies_check$Subject_ID)) |>
111   dplyr::select(
112     "Subject_ID",
113     "Study_ID",
114     "Enrolled_Date_Time",
115     "New_Intervention_Name",
116     "PI_ID",
117     "AE_Grade_3_Plus",
118     "New_Intervention_Name",
119     "eligible",
120     "Age_40",
121     "Age_65",
122     "Baseline_AE",
123     "New_Int_Name",
124     "Phase",
125     "Randomized",
126     "Combination",
127     "Sponsor_Type",
128     "Disease_Site_Group"
129   )
130
131
132 own_check <- own |>
133   dplyr::select(Subject_ID, Study_ID,
134     New_Intervention_Name, PI_ID) |>
135   dplyr::filter(Subject_ID %in%
136     unique(own_subj_mult_studies_check$Subject_ID)) |>
137   dplyr::distinct(Subject_ID, Study_ID,
138     New_Intervention_Name, PI_ID) |>
139   dplyr::group_by(Subject_ID, Study_ID,
140     New_Intervention_Name, PI_ID)

```

```

140   ###have to do New_Intervention_Name in here for correct
      department;
141 linkedDataStudies_0 <- own |>
142   dplyr::distinct(New_Intervention_Name, Study_ID,
      Subject_ID, .keep_all = TRUE) |>
143   dplyr::group_by(New_Intervention_Name, Study_ID) |>
144   dplyr::count(name = "Num_Patients")
145
146
147 linkedDataStudies <- own |>
148   dplyr::select(New_Intervention_Name) |>
149   dplyr::group_by(New_Intervention_Name)
150
151 linkedDataPIs_0 <- own |>
152   dplyr::distinct(PI_ID, New_Intervention_Name, Study_ID,
      Subject_ID) |>
153   dplyr::group_by(PI_ID, New_Intervention_Name, Study_ID) |>
154   dplyr::count(name = "Num_Patients")
155
156
157 linkedDataPIs <- own |>
158   dplyr::distinct(PI_ID, New_Intervention_Name, Study_ID) |>
159   dplyr::group_by(PI_ID)
160
161 linkedDataSubject_ID <- own |>
162   dplyr::select(
163     Subject_ID,
164     Enrolled_Date_Time,
165     Study_ID,
166     New_Intervention_Name,
167     New_Int_Name,
168     eligible,
169     Combination,
170     Randomized,
171     AE_Grade_3_Plus,
172     Age_65
173   ) |>
174   dplyr::arrange(
175     Subject_ID,
176     Enrolled_Date_Time,
177     Study_ID,
178     New_Intervention_Name,
179     New_Int_Name,
180     eligible,
181     desc(AE_Grade_3_Plus)
182   ) |>
183   dplyr::group_by(Subject_ID, Study_ID,
      New_Intervention_Name) |>
184   dplyr::filter(row_number() == 1)
185

```

```

186 reach2=function(x){
187   r=vector(length=vcount(x))
188   for (i in 1:vcount(x)){
189     n=neighborhood(x,2,nodes=i)
190     ni=unlist(n)
191     l=length(ni)
192     r[i]=(l)/vcount(x)}
193   r}
194
195 reach3=function(x){
196   r=vector(length=vcount(x))
197   for (i in 1:vcount(x)){
198     n=neighborhood(x,3,nodes=i)
199     ni=unlist(n)
200     l=length(ni)
201     r[i]=(l)/vcount(x)}
202   r}
203
204 dwreach=function(x){
205   distances=shortest.paths(x) #create matrix of geodesic
      distances
206   diag(distances)=1 # replace the diagonal with 1s
207   weights=1/distances # take the reciprocal of distances
208   apply(weights,1,sum) # sum for each node (row)
209 }
210
211 #
212
213 fpntable <- table(own$Subject_ID);
214 otable <- table(own$PI_ID);
215 rtable <- table(own$New_Intervention_Name);
216 ownSmall <- own;
217 three_way_count <- ownSmall |>
218   dplyr::select(PI_ID, New_Intervention_Name, Study_ID,
      Subject_ID) |>
219   dplyr::group_by(PI_ID, New_Intervention_Name, Study_ID,
      Subject_ID) |>
220   dplyr::count(name="freq")
221 three_way_count <- as.data.frame(three_way_count);
222
223
224 edgelist <- cbind(three_way_count$PI_ID,
      three_way_count$New_Intervention_Name,
      three_way_count$Study_ID, three_way_count$Subject_ID,
      three_way_count$freq);
225
226 colnames(edgelist) <- c("PI_ID", "New_Intervention_Name",
      "Study_ID", "Subject_ID", "freq");
227 edgelist <- as.data.frame(edgelist);
228 linkedDataPIs_0 <- as.data.frame(linkedDataPIs_0);

```



```

229 linkedDataPIs <- as.data.frame(linkedDataPIs);
230 linkedDataStudies_0 <- as.data.frame(linkedDataStudies_0);
231 linkedDataStudies <- as.data.frame(linkedDataStudies);
232 linkedDataSubject_ID <- as.data.frame(linkedDataSubject_ID);
233
234 edgelist0 <- join_all(list(edgelist, linkedDataPIs_0,
    linkedDataStudies_0), by = c("New_Intervention_Name",
    "Study_ID"), type = "left", match = "first");
235 edgelist00 <- join_all(list(edgelist0,
    linkedDataSubject_ID), by=c("Subject_ID",
    "New_Intervention_Name", "Study_ID"), type="left", match
    = "first");
236 edgelist00_tibble <- as_tibble(edgelist00);
237
238
239 edgelist <- edgelist00[,colnames(edgelist00) %in%
    c("Subject_ID", "Study_ID", "Enrolled_Date_Time",
    "New_Intervention_Name", "PI_ID", "Num_Patients",
    "eligible", "Randomized", "Combination", "freq",
    "Department", "Enrolled_Date_Time",
    "Status_Change_Date_Time", "New_Int_Name",
    "AE_Grade_3_Plus", "Age_65")];
240
241 edgelist_count <- length(edgelist$Subject_ID);
242
243 n_pi <- length(unique(edgelist$PI_ID));
244 n_subjects <- length(unique(edgelist$Subject_ID));
245 n_studies <- length(unique(edgelist$New_Intervention_Name));
246 strat <- unique(eval(parse(text=paste("edgelist$",
    "eligible", sep=""))));
247 strat <- na.omit(strat);
248
249 edgelist <- edgelist[order(edgelist$Subject_ID,
    edgelist$Enrolled_Date_Time,
    edgelist$New_Intervention_Name),];
250 #which(is.na(eval(parse(text=paste("edgelist$", var[k],
    sep=""))))) #none, good check;
251
252 edgelist <- edgelist[,colnames(edgelist) %in%
    c("Subject_ID", "Study_ID", "New_Intervention_Name",
    "PI_ID", "Num_Patients", "freq", "eligible",
    "Randomized", "Combination", "Department",
    "Enrolled_Date_Time", "Status_Change_Date_Time",
    "New_Int_Name", "AE_Grade_3_Plus", "Age_65")];
253
254 edgelist$Subject_ID <- as.character(edgelist$Subject_ID);
255 edgelist$Study_ID <- as.character(edgelist$Study_ID);
256 edgelist$PI_ID <- as.character(edgelist$PI_ID);
257 edgelist$New_Intervention_Name <-
    as.character(edgelist$New_Intervention_Name);

```

```

258 edgelist$freq <- as.numeric(as.character(edgelist$freq));
259
260 edgelistPre_st <- edgelist
261
262 counterStrat <- length(strat)
263
264 st = strat;
265 strataCat = st;
266 edgelist <- edgelistPre_st |>
267   filter(eval(parse(text="eligible"))) == st)
268 edgelist <- as.data.frame(edgelist);
269 n_studies_strata <-
    length(unique(edgelist$New_Intervention_Name));
270
271 edgelist <- edgelist[order(edgelist$Subject_ID,
    edgelist$Enrolled_Date_Time,
    edgelist$New_Intervention_Name),];
272
273
274 ###
275 ### DO THIS FOR A SIMPLER DATASET TO CHECK CODE;
276 ###
277 edgelist <- edgelist |>
278   dplyr::arrange(Subject_ID, Enrolled_Date_Time) |>
279   dplyr::group_by(Subject_ID) |>
280   dplyr::mutate(order = row_number()) |>
281   dplyr::mutate(from = Study_ID,
282     to = Study_ID,
283     order_from = order,
284     order_to = order)
285
286 igraph0 <- edgelist |>
287   dplyr::bind_rows(edgelist) |>
288   dplyr::arrange(Subject_ID, Study_ID) |>
289   dplyr::group_by(Subject_ID, Study_ID) |>
290   dplyr::ungroup() |>
291   dplyr::group_by(Subject_ID) |>
292   dplyr::group_split() |>
293   lapply(function(x) x |>
294     dplyr::mutate(x, index = 1:nrow(x),
295       direction = ifelse(index%%2 == 1,
296         "from", "to"))) |>
297   do.call(what = rbind) |>
298   dplyr::select(Subject_ID, Study_ID, direction) |>
299   tidyr::pivot_wider(
300     id_cols = c(Subject_ID),
301     names_from = direction,
302     values_from = c(Study_ID)) |>
303   tidyr::unnest(from, .drop=TRUE) |>
304   tidyr::unnest(to, .drop=TRUE)

```

```

304
305 igraph1 <- plyr::join_all(list(igraph0,
    edgelist[,c("Subject_ID", "from", "order_from")]),
    by=c("Subject_ID", "from"), type='left');
306
307 igraph2 <- plyr::join_all(list(igraph1,
    edgelist[,c("Subject_ID", "to", "order_to")]),
    by=c("Subject_ID", "to"), type='left');
308
309 igraph2 <- igraph2 |>
310   dplyr::filter(order_from < order_to) |>
311   dplyr::arrange(Subject_ID, order_from, order_to) |>
312   dplyr::group_by(Subject_ID, from) |>
313   dplyr::filter(row_number() == 1) |>
314   dplyr::mutate(Study_ID_from = from,
315                 Study_ID_to = to)
316
317 edgelist <- edgelist |>
318   dplyr::arrange(Subject_ID, Enrolled_Date_Time) |>
319   dplyr::group_by(Subject_ID) |>
320   dplyr::mutate(order = row_number()) |>
321   dplyr::mutate(New_Intervention_Name_from =
322                 New_Intervention_Name,
323                 New_Intervention_Name_to =
324                 New_Intervention_Name)
325
326 igraph3 <- plyr::join_all(list(igraph2,
    edgelist[,c("Subject_ID", "from",
327               "New_Intervention_Name_from")]), by=c("Subject_ID",
    "from"), type='left');
328
329 igraph4 <- plyr::join_all(list(igraph3,
    edgelist[,c("Subject_ID", "to",
330               "New_Intervention_Name_to")]), by=c("Subject_ID", "to"),
    type='left');
331
332 igraph5 <- igraph4 |>
333   dplyr::mutate(Study_ID_from = from,
334                 Study_ID = to,
335                 from = New_Intervention_Name_from,
336                 to = New_Intervention_Name_to) |>
337   dplyr::select(-c("New_Intervention_Name_from",
338                    "New_Intervention_Name_to"))
339
340 igraph <- igraph5 |>
341   dplyr::mutate(from = str_wrap(from, width = 30),
342                 to = str_wrap(to, width = 30)) |>
343   tidygraph::as_tbl_graph(directed = TRUE) |>
344   igraph::as_igraph()
345
346 e <- igraph::get.edgelist(igraph, names=FALSE);

```

```

341 l <- qgraph::qgraph.layout.fruchtermanreingold(e,
      vcount=vcount(igraph),
      area=30*(vcount(igraph)^2),repulse.rad=(vcount(igraph)^2.1));
342
343 # ##### Do this for a simpler graph just before
      plotting;
344 igraph_simplified <- igraph
345 E(igraph_simplified)$weight <- 1
346 igraph_simplified <- igraph::simplify(
347   igraph_simplified,
348   remove.multiple = T,
349   remove.loops = F,
350   edge.attr.comb = list(weight = "sum", "ignore")
351 )
352 E(igraph_simplified)$label <- E(igraph_simplified)$weight
353
354
355 # FOR VISUALS IN THIS REPORT
356
357 # Figure 1
358
359 g_directed <- graph(c(1, 2, 2, 3, 3, 1), directed = TRUE)
360 g_undirected <- as.undirected(g_directed)
361
362 V(g_directed)$color <- "red"
363 V(g_undirected)$color <- "red"
364 E(g_directed)$color <- "black"
365 E(g_undirected)$color <- "black"
366 set.seed(5208)
367 par(mfrow= c(1,2),mar=c(0,0,0,0)+.1)
368 plot(g_undirected,
369       vertex.label = "",
370       edge.arrow.size = 0.5,
371       vertex.size = 20)
372 set.seed(5208)
373 plot(g_directed,
374       vertex.label = "",
375       edge.arrow.size = 0.5,
376       vertex.size = 20)
377
378 # Figure 2
379
380
381
382 par(mar=c(0,0,0,0)+1)
383 plot(
384   igraph_simplified,
385   edge.label.color = "#801818",
386   edge.label = E(igraph)$label,
387   edge.label.cex = 1,

```

```

388   edge.color = "grey",
389   edge.arrow.size = 0.3,
390   vertex.size = 5,
391   vertex.shape = "square",
392   vertex.color = "orange",
393   vertex.label = V(igraph)$name,
394   vertex.label.cex = 1.0,
395   vertex.label.dist = 1.5,
396   vertex.label.degree = pi / 2,
397   edge.curved = TRUE,
398   layout = 1
399 )
400
401 # Figure 3
402
403
404 set.seed(5208)
405 par(mfrow= c(1,1),mar=c(0,0,0,0)+.1)
406 # Create two clusters
407 cluster1 <- sample(1:10, 5, replace = FALSE)
408 cluster2 <- sample(11:20, 5, replace = FALSE)
409 # Create edges within clusters
410 edges_within_cluster1 <- t(combn(cluster1, 2))
411 edges_within_cluster2 <- t(combn(cluster2, 2))
412 # Create edge connecting the clusters
413 edge_between_clusters <- matrix(c(sample(cluster1, 1),
414   sample(cluster2, 1)), ncol = 2)
414 # Combine edges
415 edges <- rbind(edges_within_cluster1,
416   edges_within_cluster2, edge_between_clusters)
416 # Create graph
417 g <- igraph::graph_from_edgelist(edges, directed = FALSE)
418 # Calculate betweenness centrality
419 betweenness_values <- igraph::edge_betweenness(g)
420 # Get the edge with the highest betweenness
421 max_betweenness_edge <- which.max(betweenness_values)
422 # Set edge color
423 igraph::E(g)$color <- "black"
424 igraph::E(g)[max_betweenness_edge]$color <- "red"
425 g <- igraph::induced_subgraph(g, which(igraph::degree(g) >
426   0))
426 # Plot the graph
427 plot(
428   g,
429   vertex.label = "",
430   vertex.color = "grey",
431   edge.curved = FALSE,
432   edge.label = NA
433 )
434

```

```

435 # Figure 5
436
437 set.seed(5208)
438 par(mfrow= c(1,1),mar=c(0,0,0,0)+.1)
439
440 num_nodes <- 6
441
442 # Create an empty graph
443 g <- igraph::make_empty_graph(n = num_nodes)
444
445 # Add edges to connect all nodes to the central node (node
    1)
446 for (i in 2:num_nodes) {
447   g <- igraph::add_edges(g, c(1, i))
448 }
449
450 g |>
451   igraph::as.undirected() |>
452   plot(
453     vertex.label="",
454     vertex.color = ifelse(igraph::V(g)== 1, "red", "grey"),
455     edge.color = "black"
456   )
457
458
459 # Figures 7-10
460
461
462 # Putting this chunk here
463 gn_igraph <- igraph::cluster_edge_betweenness(igraph)
464
465 louvain_igraph <- igraph |>
466   igraph::as.undirected() |>
467   igraph::cluster_louvain()
468
469 sp_igraph <- igraph |>
470   ig.degree.betweenness::cluster_degree_betweenness()
471
472
473 # Figure 7
474
475 par(mar=c(0,0,0,0)+1)
476 plot(
477   gn_igraph,
478   igraph_simplified,
479   edge.label.color = "#801818",
480   edge.label = E(igraph)$label,
481   edge.label.cex = 1,
482   edge.color = "grey",
483   edge.arrow.size = 0.3,

```

```

484     vertex.size = 5,
485     vertex.shape = "square",
486     vertex.color = "orange",
487     vertex.label = V(igraph)$name,
488     vertex.label.cex = 1.0,
489     vertex.label.dist = 1.5,
490     vertex.label.degree = pi / 2,
491     edge.curved = TRUE,
492     layout = l
493 )
494
495 # Figure 8
496
497 par(mar=c(0,0,0,0)+1)
498 plot(
499     louvain_igraph,
500     igraph_simplified,
501     edge.label.color = "#801818",
502     edge.label = E(igraph)$label,
503     edge.label.cex = 1,
504     edge.color = "grey",
505     edge.arrow.size = 0.3,
506     vertex.size = 5,
507     vertex.shape = "square",
508     vertex.color = "orange",
509     vertex.label = V(igraph)$name,
510     vertex.label.cex = 1.0,
511     vertex.label.dist = 1.5,
512     vertex.label.degree = pi / 2,
513     edge.curved = TRUE,
514     layout = l
515 )
516
517 # Figure 9
518
519 par(mar=c(0,0,0,0)+1)
520 plot(
521     sp_igraph,
522     igraph_simplified,
523     edge.label.color = "#801818",
524     edge.label = E(igraph)$label,
525     edge.label.cex = 1,
526     edge.color = "grey",
527     edge.arrow.size = 0.3,
528     vertex.size = 5,
529     vertex.shape = "square",
530     vertex.color = "orange",
531     vertex.label = V(igraph)$name,
532     vertex.label.cex = 1.0,
533     vertex.label.dist = 1.5,

```

```

534     vertex.label.degree = pi / 2,
535     edge.curved = TRUE,
536     layout = l
537 )
538
539 # Figure 10
540
541 all_degree<- igraph::degree(igraph) |>
542   as.data.frame()|>
543   tibble::rownames_to_column()|>
544   dplyr::rename(degree='igraph::degree(igraph)' ,
545                 study=rowname)
546
547 in_degree <- igraph::degree(igraph, mode = "in")|>
548   as.data.frame()|>
549   tibble::rownames_to_column()|>
550   dplyr::rename(in_degree='igraph::degree(igraph, mode =
551     "in")' ,
552                 study=rowname)
553
554 out_degree <- igraph::degree(igraph, mode = "out") |>
555   as.data.frame()|>
556   tibble::rownames_to_column()|>
557   dplyr::rename(out_degree='igraph::degree(igraph, mode =
558     "out")' ,
559                 study=rowname)
560
561 degree_df <- merge(in_degree,
562                   out_degree)|>
563   merge(all_degree)|>
564   dplyr::mutate(in_degree = -in_degree)|>
565   tidyr::pivot_longer(cols = c(in_degree,out_degree))
566
567 ggplot(degree_df,
568       mapping = aes(y =reorder(study, degree), x = -value,
569                       fill = name))+
570   theme_minimal()+
571   geom_col()+
572   geom_hline(yintercept = 2.5,linetype='dashed',lwd=1)+
573   geom_hline(yintercept = 12.5,linetype='dashed',lwd=1)+
574   theme(axis.title.y = element_blank(),
575         legend.title = element_blank(),
576         legend.position = "bottom",
577         axis.title.x = element_blank())+
578   scale_fill_manual(labels = c("Referrals In", "Referrals
579     Out"), values = scales::hue_pal()(2))+
580   scale_x_continuous(labels = abs)
581
582 # Tables

```



```

580 # Table 1
581
582 gn_df <- data.frame(
583   Intervention = igraph::V(igraph)$name,
584   "Patient Refferalls: In" =
585     igraph::degree(igraph,mode="in"),
586   "Patient Referrals: Out" = igraph::degree(igraph,
587     mode="out"),
588   "Total Patient Refferalls" = igraph::degree(igraph,
589     mode="total"),
590   row.names = NULL,
591   check.names = FALSE
592 ) |>
593 dplyr::group_by(Intervention) |>
594 dplyr::summarise(
595   'Refferalls In' = sum('Patient Refferalls: In'),
596   'Referrals Out' = sum('Patient Referrals: Out'),
597   'Total' = sum('Total Patient Refferalls')
598 )
599
600 gt::gt(gn_df)|>
601 gt::tab_header("Table 1: Girvan-Newman communities
602   identified. Each intervention is their own
603   community.")|>
604 gt::cols_width(
605   Intervention ~ gt::pct(40),
606   'Refferalls In' ~ gt::pct(15),
607   'Referrals Out' ~ gt::pct(20),
608   'Total' ~ gt::pct(15)
609 ) |>
610 gt::tab_options(table.font.size=42)
611
612 # Table 2
613
614 louvain_df <- data.frame(
615   Intervention = igraph::V(igraph)$name,
616   Community = paste0("Community: ",
617     igraph::membership(louvain_igraph)|> as.vector()),
618   "Patient Refferalls: In" =
619     igraph::degree(igraph,mode="in"),
620   "Patient Referrals: Out" = igraph::degree(igraph,
621     mode="out"),
622   "Total Patient Refferalls" = igraph::degree(igraph,
623     mode="total"),
624   row.names = NULL,
625   check.names = FALSE
626 )

```

```

621 | louvain_df |>
622 |   dplyr::group_by(Community, Intervention) |>
623 |   dplyr::summarise(
624 |     'Refferalls In' = sum('Patient Refferalls: In'),
625 |     'Referrals Out' = sum('Patient Referrals: Out'),
626 |     'Total' = sum('Total Patient Refferalls')
627 |   )|>
628 |   gt::gt()|>
629 |   gt::tab_header("Table 2: Louvain communities identified
630 |     and grouped interventions.")|>
631 |   gt::cols_width(
632 |     Intervention ~ gt::pct(40),
633 |     'Refferalls In' ~ gt::pct(15),
634 |     'Referrals Out' ~ gt::pct(20),
635 |     'Total' ~ gt::pct(15)
636 |   )|>
637 |   gt::tab_options(table.font.size=42)
638 |
639 |
640 | # Table 3
641 |
642 | sp_df <- data.frame(
643 |   Intervention = igraph::V(igraph)$name,
644 |   Community = paste0("Community: ",
645 |     igraph::membership(sp_igraph)|> as.vector()),
646 |   "Patient Refferalls: In" =
647 |     igraph::degree(igraph, mode="in"),
648 |   "Patient Referrals: Out" = igraph::degree(igraph,
649 |     mode="out"),
650 |   "Total Patient Refferalls" = igraph::degree(igraph,
651 |     mode="total"),
652 |   row.names = NULL,
653 |   check.names = FALSE
654 | )
655 |
656 | sp_df |>
657 |   dplyr::group_by(Community, Intervention) |>
658 |   dplyr::summarise(
659 |     'Refferalls In' = sum('Patient Refferalls: In'),
660 |     'Referrals Out' = sum('Patient Referrals: Out'),
661 |     'Total' = sum('Total Patient Refferalls')
662 |   )|>
663 |   gt::gt()|>
664 |   gt::tab_header("Table 3: Smith-Pittman communities and
665 |     identified and grouped interventions.")|>
666 |   gt::cols_width(
667 |     Intervention ~ gt::pct(40),
668 |     'Refferalls In' ~ gt::pct(15),
669 |     'Referrals Out' ~ gt::pct(20),

```

```
665     'Total' ~ gt::pct(15)
666   )
```