Understanding Patterns in Patient Discharge Summaries using Network Analysis

Cody Dunne

I. INTRODUCTION

This paper describes an exploration of medical patients and concepts related to their care. These concepts have been extracted from the patient discharge summaries, and include any associated symptoms, diseases, drugs, and procedures. They were provided by Todd Johnson, director of Biomedical Informatics at the University of Kentucky and analyzed with the help of Dr. Seth Powsner from Yale Medical. The goal in analyzing this dataset was to see if motif simplification would help medical researchers understand overall patient trends, such as comparing the efficacy of competing treatments for the same condition.

II. EXPLORATION ENVIRONMENT

For these explorations we used the NodeXL network analysis tool [Smi+09; Smi+10; HSS11], a free and open source add-in for Excel 2007/2010/2013. NodeXL, shown in Fig. 1, is tailored to provide powerful features while still being easy to learn. The Excel integration allows rapid data processing using standard formulas and macros, but NodeXL also provides calculators for network statistics, automatic layout algorithms, visual attribute encodings, dynamic filters, direct manipulation, coordinated views, and importers from online social networks and common network file formats like GraphML, Pajek, and UCINET. These importers are especially important for helping novice users collect datasets that are of interest to them like Twitter keyword searches, their Facebook network, or their personal email collection. Once the analysis is completed, users can upload the resulting visualizations and data to the NodeXL Graph Gallery [Smi+13] to share with others.

NodeXL is widely used in many disciplines and has a full-time developer as well as a team of volunteer advisors and developers. Over 25 introductory courses on network analysis have used NodeXL and its companion book [HSS11] as part of their curriculum,[1] due mainly to its ease of use, open source nature, and design focus on novice users.

The techniques discussed in this paper are implemented and made publicly available in NodeXL. The motif simplification approach detailed in Section III is currently shipping in NodeXL for anyone to use and build upon. Of the Group-in-a-Box layouts (Section IV), the Treemap GIB layout is already available in NodeXL. The Croissant-Donut and Force-Directed variants have been implemented and will be pushed to the trunk shortly.

A. *NodeXL Interface*

The basic interface of NodeXL is shown in Fig. 1. The left side provides several worksheets in an Excel workbook that represents the network: one each for the nodes, edges, groups, group members, and overall metrics. Each worksheet has several columns, including basic information about the network like the nodes and edges between them. Additionally, there are places to insert columns for node or edge attributes and calculated metrics, as well as columns that control the visual display of each network item. These include color, shape, size, label, tooltip, display position, and the like. Any of these visual properties can be automatically filled based on the metric or attribute columns using a special autofill dialog. Moreover, standard Excel formulas or macros can be used for arbitrary calculations and scales within the tool. The Excel ribbon is customized with a new tab for many of the common operations users perform on networks, including the autofill feature.

The visualization pane shown in the right of Fig. 1 displays a node-link visualization based on the network in the workbook. Whenever the contents of the workbook is updated, the visualization pane can be refreshed using a button. The pane also provides users with several automatic layout algorithms to arrange the network, and any automatic or manual adjustments to the node positions are stored in the workbook as well. Moreover, the contents of the visualization can be filtered using a dynamic filters dialog. Additional windows can be opened for filtering the visible network, autofilling visual property columns based on metrics or attributes, and running automated analyses of several networks sequentially.

The worksheet view and the visualization pane are connected using brushing, where any selection in one is reflected in the other. Clicking a node in the visualization or dragging a box around several causes the associated rows to be selected in the nodes worksheet. Likewise, any incident edges are selected in the edges worksheet. The reverse is also true. Any nodes or edges selected in the worksheets are highlighted in the visualization pane as well.

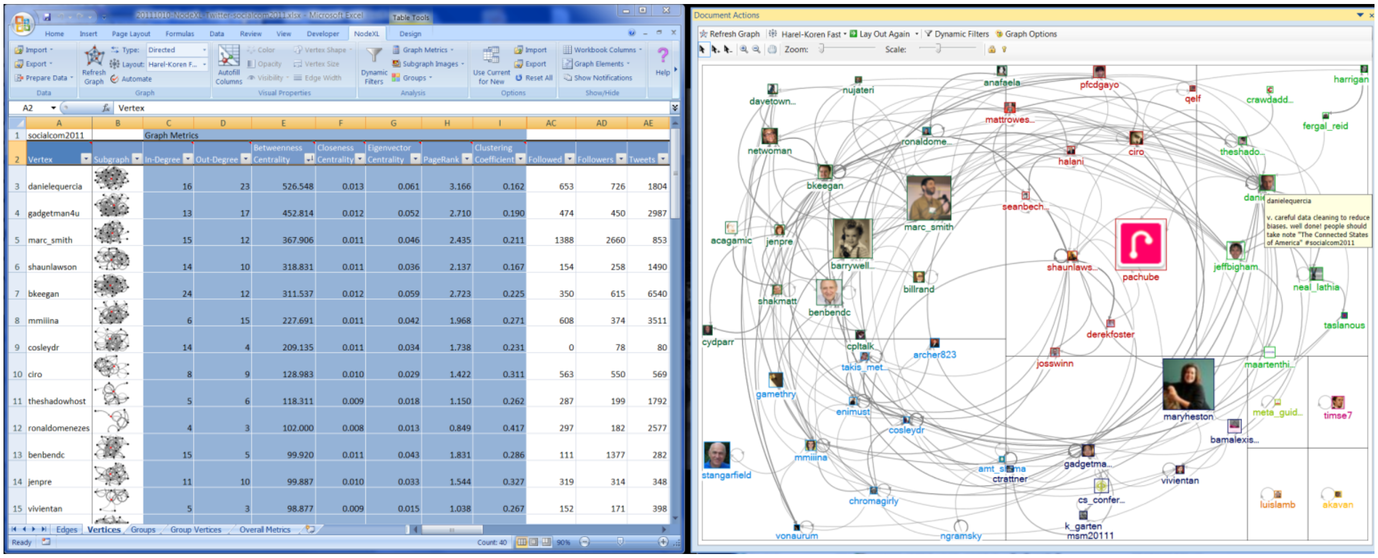
III. MOTIF SIMPLIFICATION

A. *Technique*

Many complex networks are littered with recurring topologic patterns or **motifs**, either because of the network structure or data collection methods. Three of these motifs are shown in the top row of Fig. 2. Regardless of their cause, some frequently expressed motifs contain little information compared to the space they occupy in the visualization. We

---

[1] nodexl.codeplex.com/wikipage?title=NodeXL%20Teaching%20Resources



	A	B	C	D	E	F	G	H	I	AC	AD	AE
1	socialcom2011		Graph Metrics									
2	Vertex	Subgraph	In-Degree	Out-Degree	Betweenness Centrality	Closeness Centrality	Eigenvector Centrality	PageRank	Clustering Coefficient	Followed	Followers	Tweets
3	danielequercia		16	23	526.548	0.013	0.061	3.166	0.162	653	726	1804
4	gadgetman4u		13	17	452.814	0.012	0.052	2.710	0.190	474	450	2987
5	marc_smith		15	12	367.906	0.011	0.046	2.435	0.211	1388	2660	853
6	shaunlawson		14	10	318.831	0.011	0.036	2.137	0.167	154	258	1490
7	bkeegan		24	12	311.537	0.012	0.059	2.723	0.225	350	615	6540
8	mmiiina		6	15	227.691	0.011	0.042	1.968	0.271	608	374	3511
9	cosleydr		14	4	209.135	0.011	0.034	1.738	0.231	0	78	80
10	ciro		8	9	128.983	0.010	0.029	1.422	0.311	563	550	569
11	theshadowhost		5	6	118.311	0.009	0.018	1.150	0.262	287	199	1792
12	ronaldomenezes		4	3	102.000	0.008	0.013	0.849	0.417	297	182	2577
13	benbendc		15	5	99.920	0.011	0.043	1.831	0.286	111	1377	282
14	jenpre		11	10	99.887	0.010	0.033	1.544	0.327	319	314	348
15	vivientan		5	3	98.877	0.009	0.015	1.038	0.267	152	171	398

danielequercia

v. careful data cleaning to reduce biases. well done! people should take note "The Connected States of America" #socialcom2011

Fig. 1: The NodeXL [Smi+10] workspace. The dual pane view of network data and metrics (left pane) with node-link visualization (right pane) provide an integrated snapshot of statistics and visualization, along with built-in functions and controls that support exploration and discovery. Individual worksheets separate network analysis tasks into separate categories, closely aligned with topology and attribute-based tasks, such as “Edges”, “Vertices” (nodes), and “Groups.”

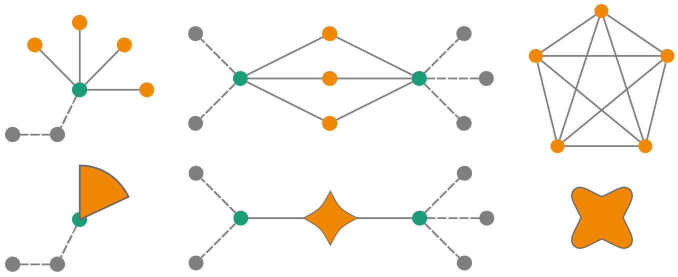


Fig. 2: Fan, connector, and clique motifs (top) and their glyphs (bottom).

use a new technique called **motif simplification** [DS13; SD12; Dun13], in which common repeating motifs are replaced with compact yet meaningful glyphs. We simplify the three frequently occurring and high-payoff motifs shown in Fig. 2: **fans** of nodes with a single neighbor, **connectors** that link a set of anchor nodes, and **cliques** of completely connected nodes.

User studies showed [DS13; Dun13] that simplifying these motifs (1) reduces screen space used and layout effort, (2) can reveal hidden relationships, and (3) is quite beneficial for many network analysis tasks both in the time users took and their accuracy/error. Unlike other approaches, motif simplification is able to achieve these benefits while maintaining user awareness of the underlying topology.

B. Initial Exploration

Dr. Johnson suggested that we investigate two medication concepts in the anonymized network, “hops5325” and “orch7323”, where “hops” stands for Hazardous or Poisonous Substance and “orch” indicates Organic Chemical. We extracted from the overall network only those patients connected to “hops5325” and/or “orch7323”, as well as any additional concepts associated with those patients (a 2-degree subnetwork). This resulted in 433 patients connected to 4701 concepts, including “hops5325” and “orch7323”. Fig. 3 shows a node-link visualization of this subnetwork using the Harel-Koren FMS layout [HK02]. The two ego concepts “hops5325” and “orch7323” are shown large and in orange, other concepts are blue, and the patients are purple triangles. This initial view does not show much structure, aside from “orch7323” being more central to the network and connected to more of the patients. Applying motif simplification, specifically the fan and connector motifs, reduces the complexity somewhat but not spectacularly (Fig. 4). The exact reduction is from 5134 nodes to 2695 nodes and 439 motif glyphs, and from 31,518 edges to 28,375 edges and meta-edges.

Now that we have the motifs, we can use them to highlight or drill down into interesting patterns. Fig. 5 shows the largest fan motifs highlighted in red, where each fan has at least 20 concepts and up to 42 for the largest. These concepts are unique to a single patient, and the patients and their connections to the fans are highlighted in red as well. A medical researcher may be interested in exploring these singleton concept groups and drilling down to them or, alternatively, filtering them out to see the more common patterns. In this case we drill down to show only those patients and their connected concepts, displayed in Fig. 6 without simplification. “hops5325” is peripheral to this network, only connected to two patients on the right. In the simplified view (Fig. 7), “hops5325” is in a connector motif with three other concepts that are only connected to those two patients: “orch7268”, “hlca5025”, and “hlca5238”. Interestingly, only one of these patients is connected to “orch7323”. Another pattern of note is the large connector motif on the left, which consists of 36 concepts associated with two other patients who are connected

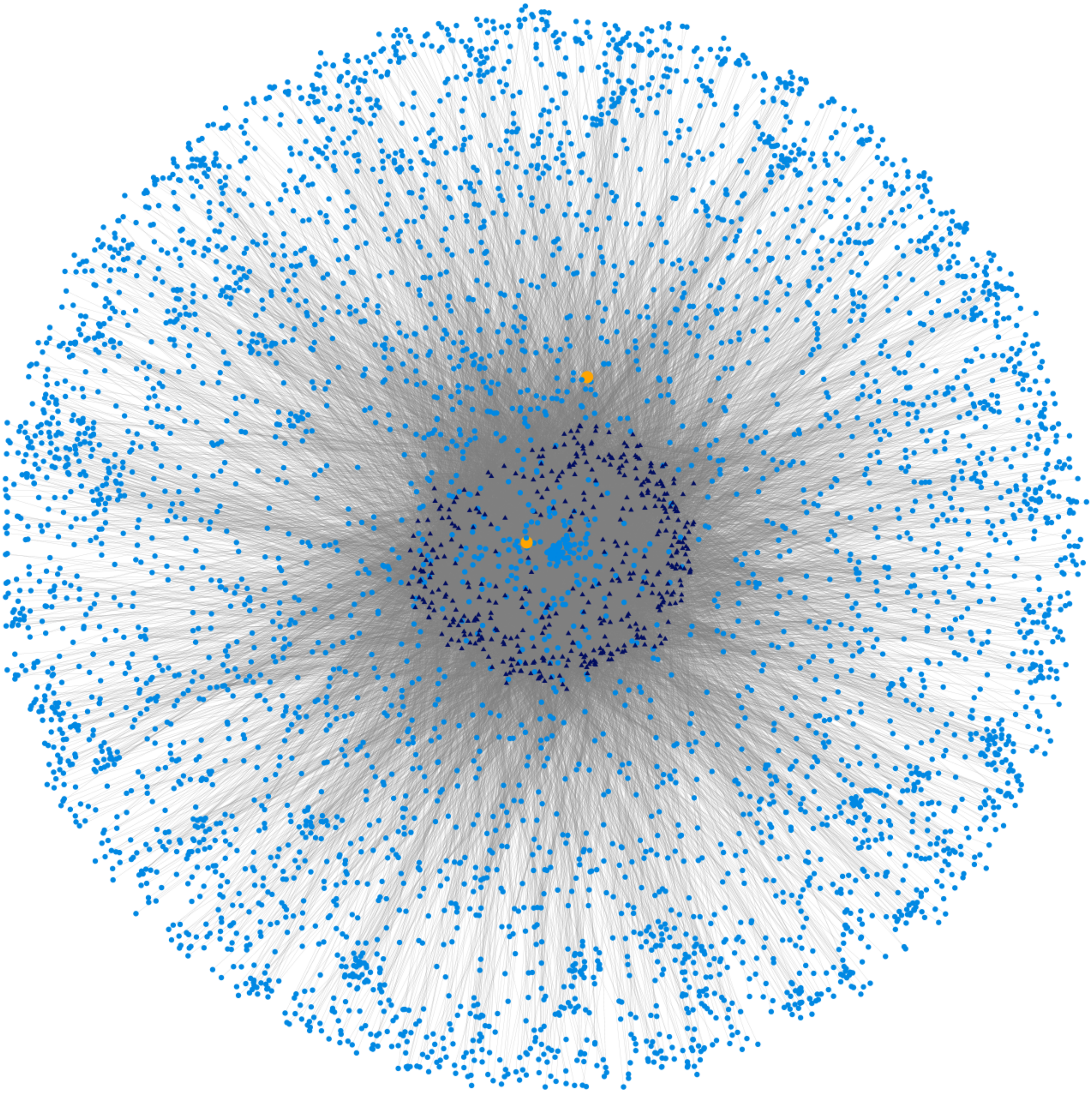


Fig. 3: Patients related to concepts from their medical discharge reports. This subnetwork focuses on the concepts “hops5325” and “orch7323” (orange discs) and their associated patients (purple triangles) and concepts (blue discs). The network is laid out using the Harel-Koren FMS layout algorithm [HK02].

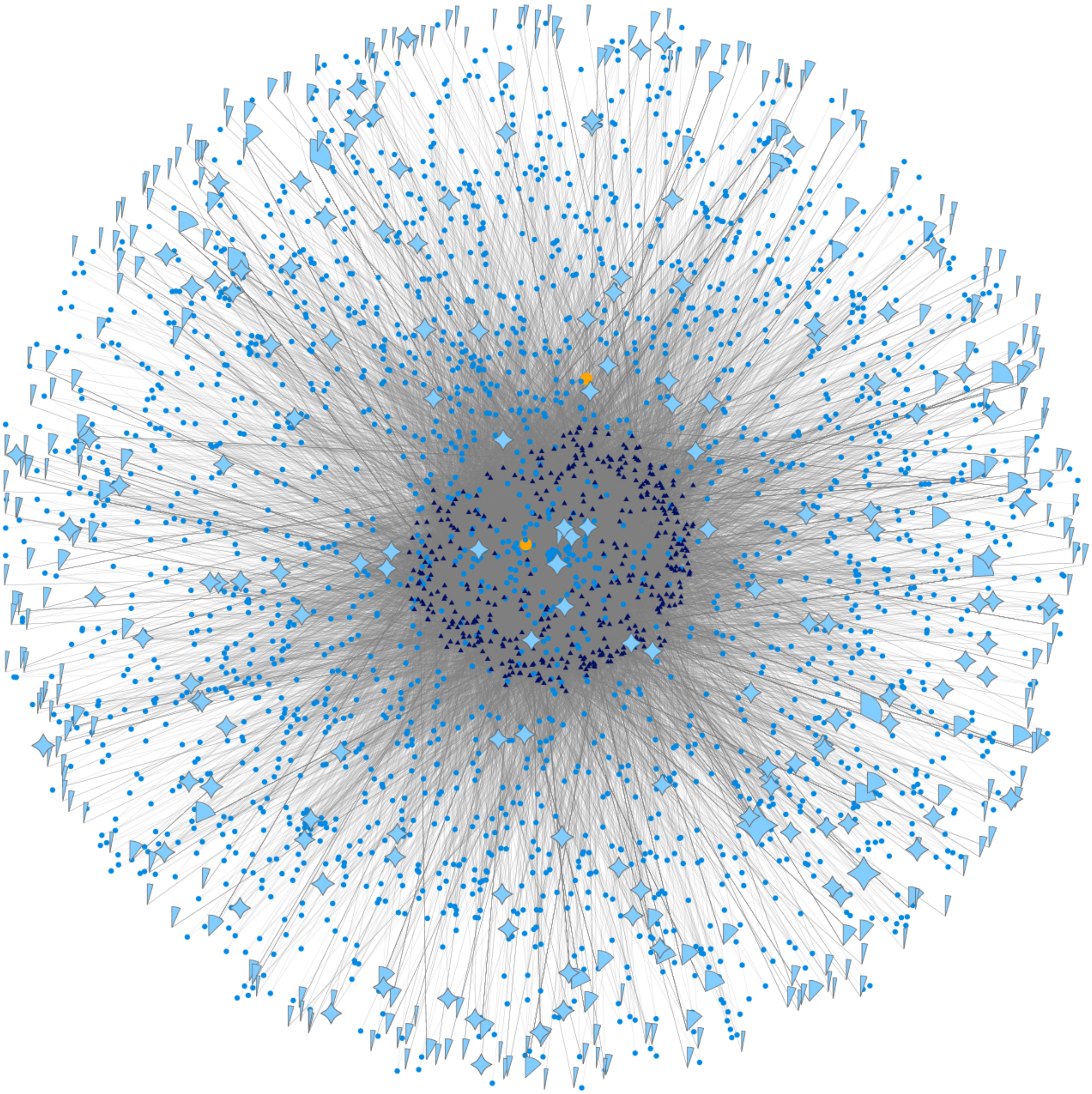


Fig. 4: Patients and concepts from Fig. 3 after applying fan and connector motif simplification.

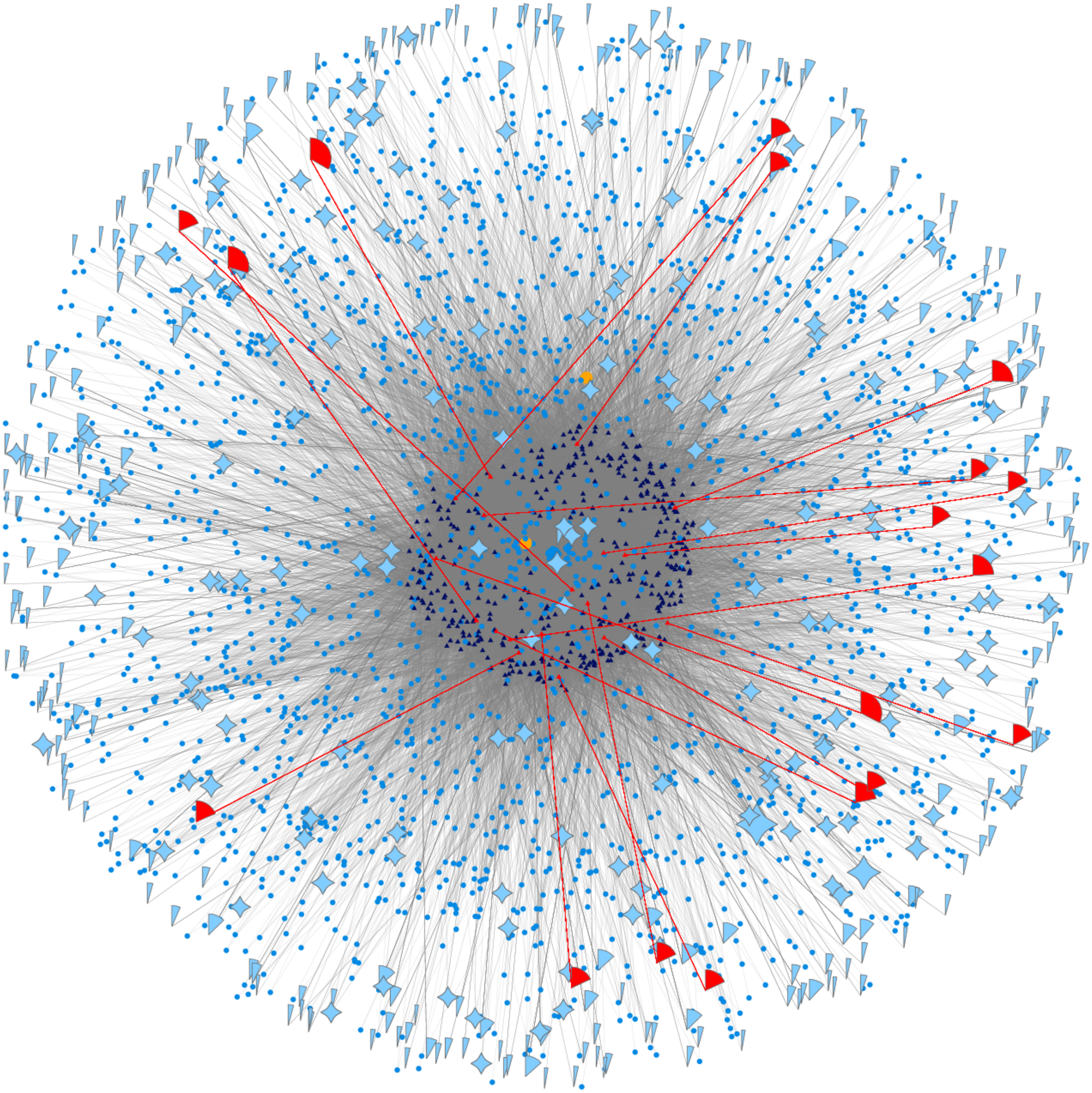


Fig. 5: Simplified patient and concept network from Fig. 4 with fans of 20 or more concepts highlighted. This shows groups of concepts that are uniquely associated with a single patient. Edges from these fans to their associated patient, as well as the patient themselves, are highlighted too.

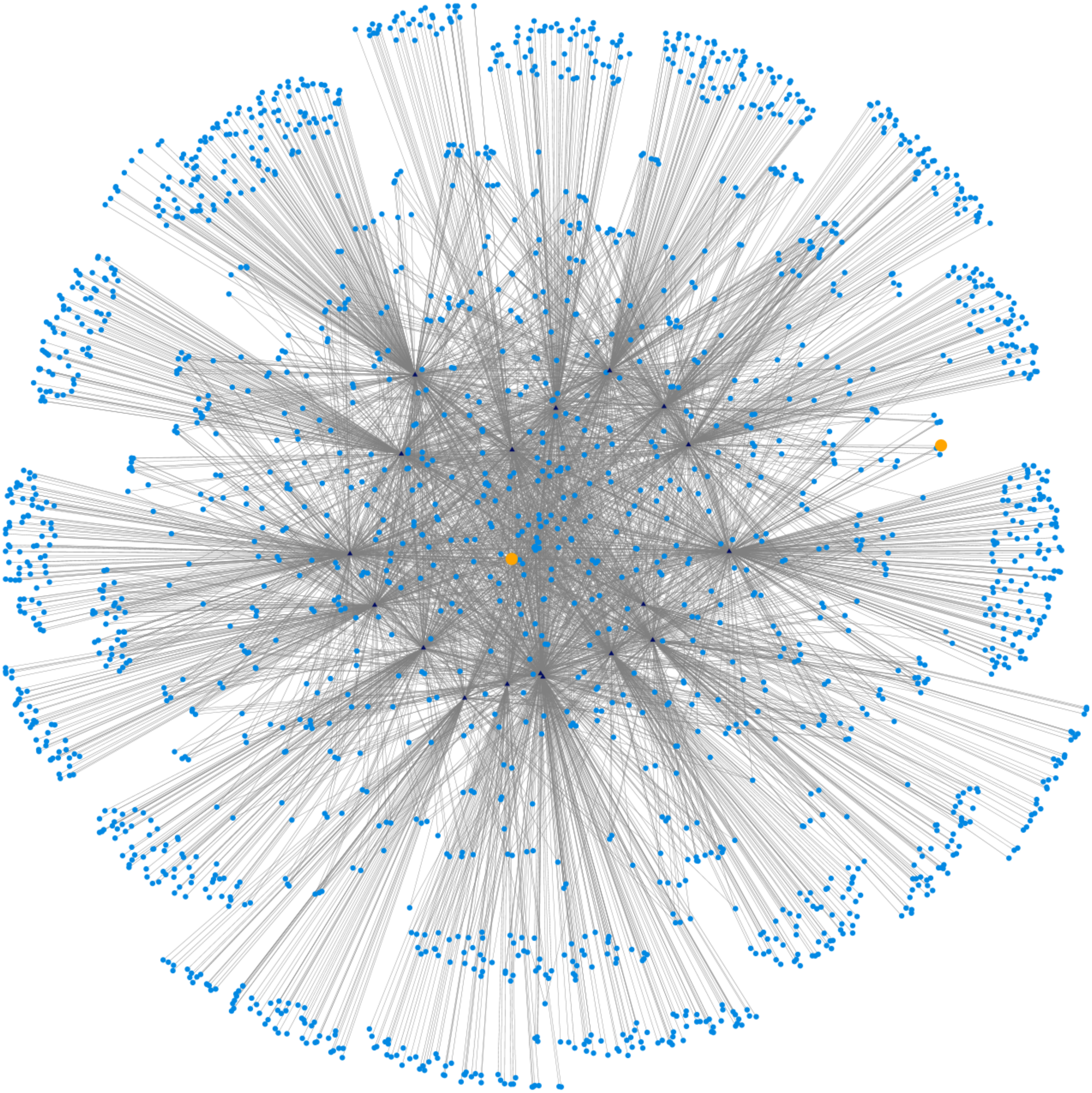


Fig. 6: Patient and concept network of only the patients connected to the large highlighted fans from Fig. 5, as well as any associated concepts. The initial “hops5325” concept is on the far right, connected to only two patients.

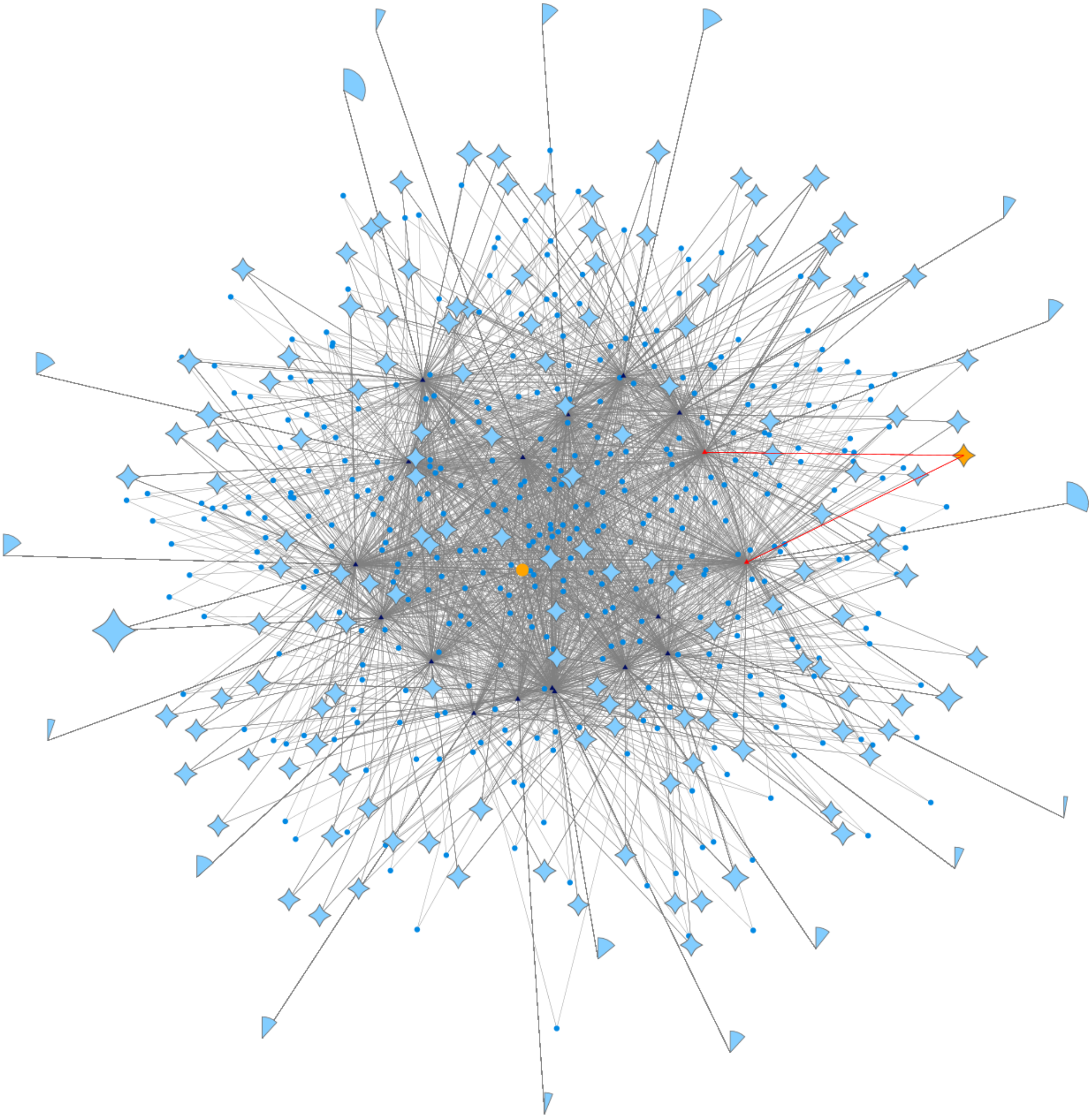


Fig. 7: Patient and concept network from Fig. 6 after applying motif simplification. The connector motif which contains the initial “hops5325” concept and three other concepts is highlighted in orange. These four concepts are only connected to two patients.

to “orch7323”. These concepts are “aapp155”; “dsyn 2382, 2732, 2842, 3006, 3092, 3171, 3464, 3576, 3577, 3817, 3837, 3927, 4009, 4261, 4528, and 4827”; “lbpr 5981, 5990, and 6419”; “mobd 6668, 6673, 6688, 6690, and 6715”; “orch 7921, 8368, and 8369”; “patf 8787, 8818, and 8983”; “phsu9097”; and “topp 10357, 10429, and 10856”.

An alternate kind of exploration is visible in Fig. 8, where we have highlighted connectors of concepts connected to at least 20 patients. These small connectors consist of two or more concepts that occur with many patients in the exact same way, but the connectors each have different sets of the 433 patients as anchors.

The true power of motif simplification becomes evident when we drill down to only show the patients connected to four specific concepts. We chose our original “hops3525” and “orch7323”, as well as two other Hazardous or Poisonous Substances: “hops5323” and “hops5324”. The node-link visualization of these relationships is partially understandable (Fig. 9), but after applying motif simplification the aggregate patient relationships between the concepts are much more clear (Fig. 9). Note that here the motifs consist of patients, not concepts. It is immediately visible that two patients are connected to all four concepts and one patient is shared between only the “hops” concepts. Another 7 patients connect “orch7323” and “hops5325” while 67 connect “hops5323”, “hops5324”, and “orch7323”. Of course, 339 patients only have “orch7323” as a concept while only 17 are only connected to “hops5325”.

Overall, we believe that motif simplification can definitely help medical researchers understand the relationships between patients and a small number of concepts, as in Fig. 10. For larger datasets with thousands of concepts, the motifs seem to highlight particularly unusual connections like large groups of concepts associated with one patient or a few patients. To understand these relationships in detail, the motifs can be used to drill down to the relevant parts of the network.

C. Second Exploration

For the second exploration of this space Dr. Powsner, Dr. Johnson, and I looked for a particular condition. Specifically, we searched for lung dysfunction causing atrial size increase, followed by worsening pacing, and which leads to atrial fibrillation, thrombosis, or stroke. We searched the database for concepts matching “atrial”, “stroke”, “thromb”, or “hypertension” and culled these related terms from the results:

- “atrial”
  - left atrial dilatation
  - atrial dilatation
  - atrial premature complexes
  - chronic atrial fibrillation
  - left atrial hypertrophy
  - paroxysmal atrial fibrillation
  - rapid atrial fibrillation
  - recurrent atrial fibrillation
  - tachycardia ectopic atrial
  - atrial arrhythmia
  - atrial fibrillation
  - atrial flutter
  - atrial tachycardia
  - ectopic atrial pacemaker
  - paroxysmal atrial flutter
  - permanent atrial fibrillation
  - persistent atrial fibrillation
- “stroke”
  - embolic stroke
  - ischemic stroke
  - multiple strokes
  - paralytic stroke
- “thromb”
  - non-occlusive thrombus
  - antithrombin iii deficiency
  - cerebral thrombosis
  - deep vein thrombosis
  - deep vein thrombosis of lower limb
  - lateral sinus thrombosis
  - left ventricular thrombus
  - mural thrombus
  - mural thrombus of heart
  - retinal vein thrombosis
  - splenic vein thrombosis
  - thrombocytosis
  - thromboembolism
  - thrombophlebitis
  - thrombophlebitis of upper extremities
  - thrombosis of cerebral veins
  - thrombosis of mesenteric vein
  - thrombosis of renal vein
  - thrombus
  - cerebral arterial thrombosis
  - jugular vein thrombosis
  - thrombosis
  - thrombosis of venous sinuses
  - venous thromboembolism
  - venous thrombosis
  - thrombectomy
  - thrombolytic therapy
- “hypertension”
  - benign essential hypertension
  - benign hypertension
  - borderline hypertension
  - diastolic hypertension
  - essential hypertension
  - labile hypertension
  - malignant essential hypertension
  - malignant hypertension
  - prehypertension
  - reactive hypertension
  - secondary hypertension nos
  - systolic hypertension
  - transient hypertension
  - uncontrolled hypertension
  - hypertension orthostatic
  - postoperative hypertension
  - hypertension medication

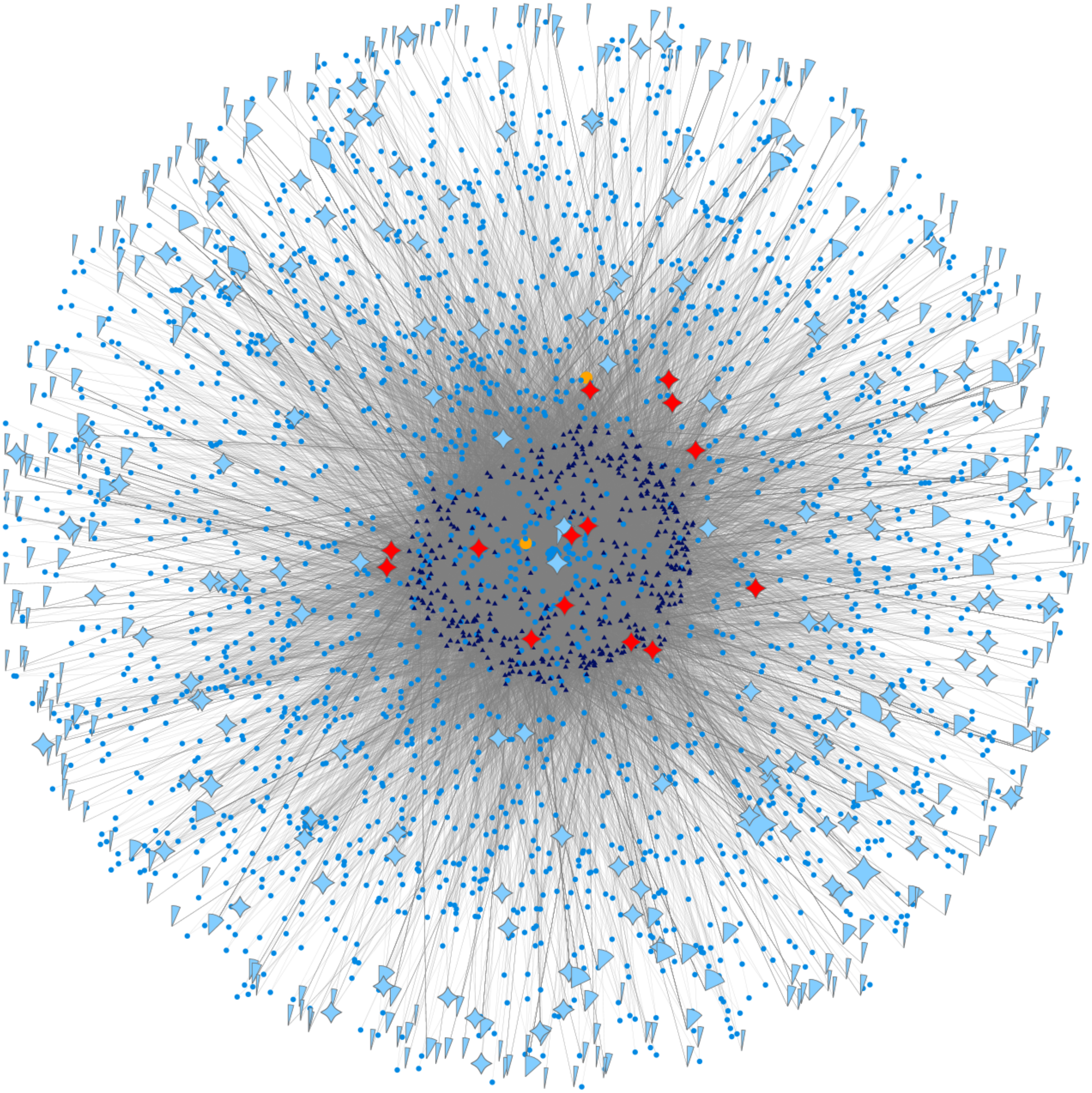


Fig. 8: Patients and concepts from the original simplified view in Fig. 4. Connector motifs of concepts connected to at least 20 patients are highlighted.

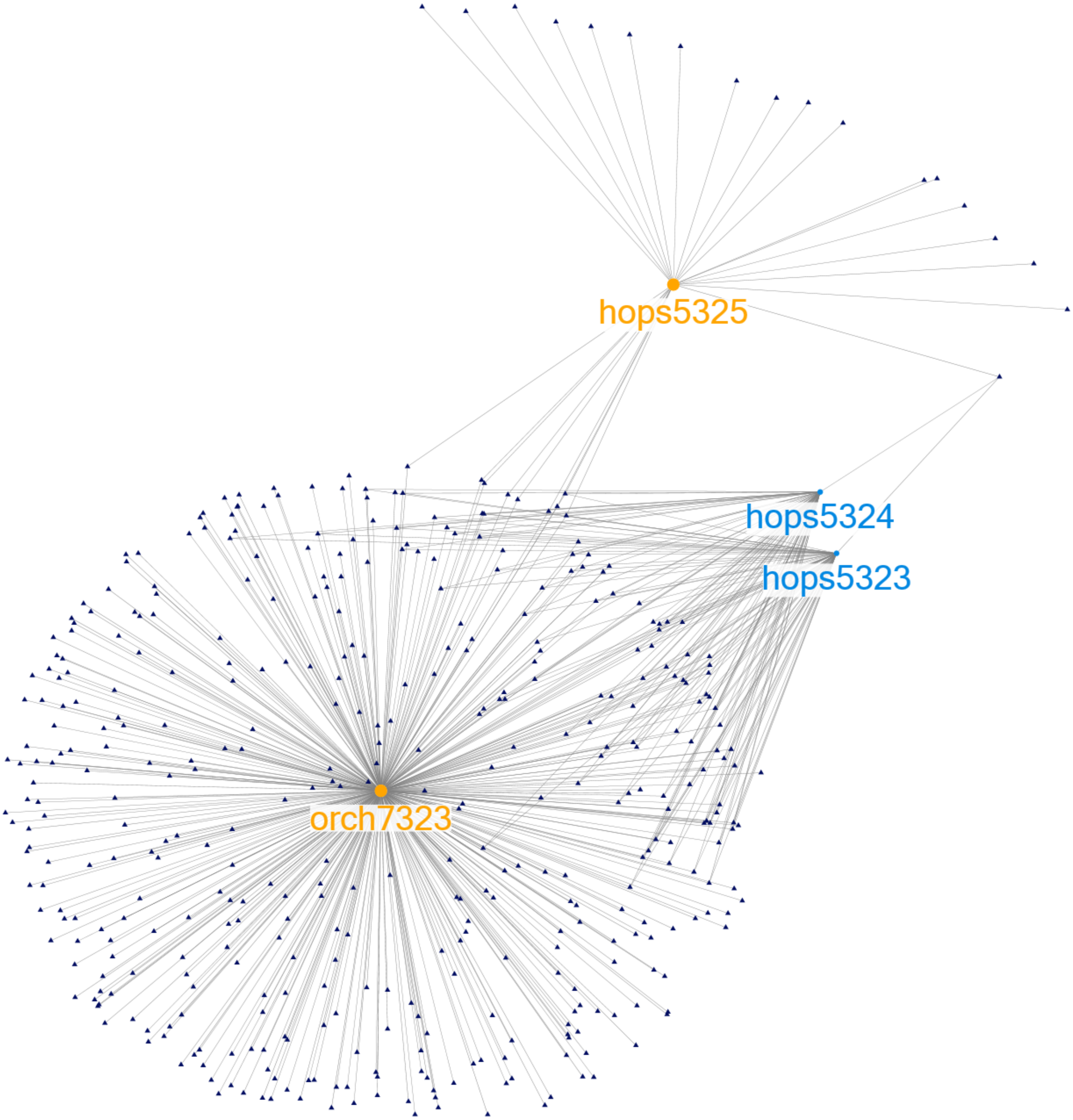


Fig. 9: Patients and concepts from Fig. 3, after drilling down to only those patients connected to our original “hops3525” and “orch7323”, as well as two other Hazardous or Poisonous Substances: “hops5323” and “hops5324”.

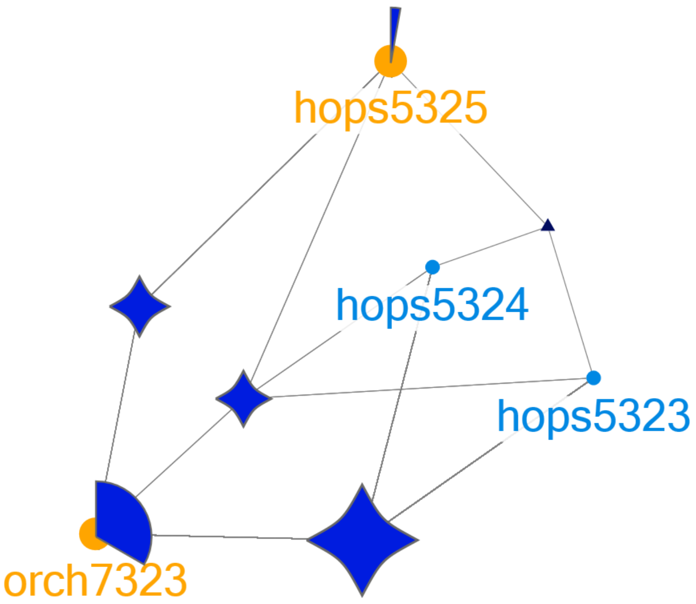


Fig. 10: A simplified view of the patients and concepts in Fig. 9, which highlights the aggregate patient relationships between the concepts.

The connections between these concepts and their associated patients is shown in the visualization of Fig. 11. The visualization in this figure uses motif simplification to replace repeating patterns of patients connecting concepts with representative glyphs. However, in three cases a pair of concepts only related to one patient were replaced with a fan motif glyph. A few of the motifs stand out from the image. First, the large 2-connector motif on the far left contains 450 patients, which are connected only to "tobacco" and "tobacco preparation". Next, the massive fan motif on the far right contains the 171 patients that were only connected to "asprin".

1) Tobacco and Tobacco Preparation Subgraphs: Using NodeXL's Select Subgraphs feature, we selected all the concepts connected via an intermediary patient to the "tobacco" and "tobacco preparation" concepts. Here is the resulting list of concepts:

- aspirin
- aspirin 325 mg
- aspirin 80 mg
- aspirin 81 mg
- atrial dilatation
- atrial fibrillation
- atrial flutter
- atrial premature complexes
- atrial tachycardia
- bayer aspirin
- benign essential hypertension
- borderline hypertension
- chronic atrial fibrillation
- coumadin
- deep vein thrombosis
- deep vein thrombosis of lower limb
- essential hypertension
- heparin
- ischemic stroke
- labile hypertension
- left atrial dilatation
- lovenox
- multiple strokes
- paroxysmal atrial fibrillation
- permanent atrial fibrillation
- plavix
- thrombocytosis
- thrombolytic therapy
- thrombosis
- ticlid
- ticlopidine 250 mg
- tobacco
- tobacco preparation
- tobacco use disorder
- uncontrolled hypertension
- venous thrombosis
- warfarin
- warfarin 5 mg
- warfarin sodium

We analyzed these 39 concepts, and the 58 others that were not connected, to see whether the network structure matches known relationships. The outcome of our analysis was that these network techniques help pull together the majority of the terms that should be related, but slight variations in language can be problematic. For example, each of these terms were not connected but were attributed to local nomenclature differences: "aspirin prophylaxis", "buffered aspirin", "clopidogrel", and "oral form aspirin".

2) Atrial Fibrillation and Hypertension: We then examined the subset of concepts that contained "atrial fibrillation" or "hypertension". Fig. 12 shows the resulting visualization. A few things stand out from this now much simpler network. First, we see that "chronic atrial fibrillation", "paroxysmal atrial fibrillation", and "atrial fibrillation" all show up in

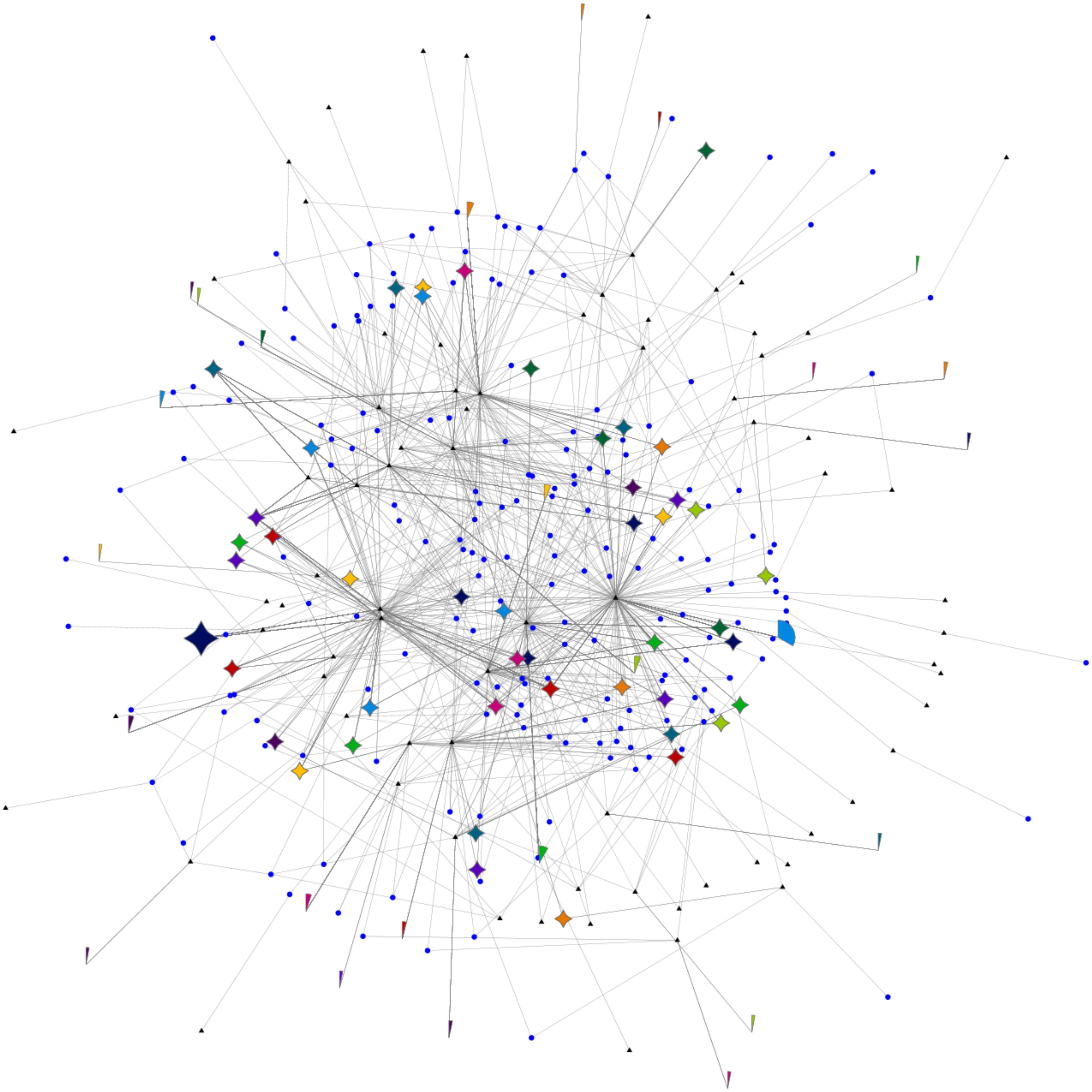


Fig. 11: Network of concepts matching “atrial”, “stroke”, “thromb”, or “hypertension” and their associated patients.

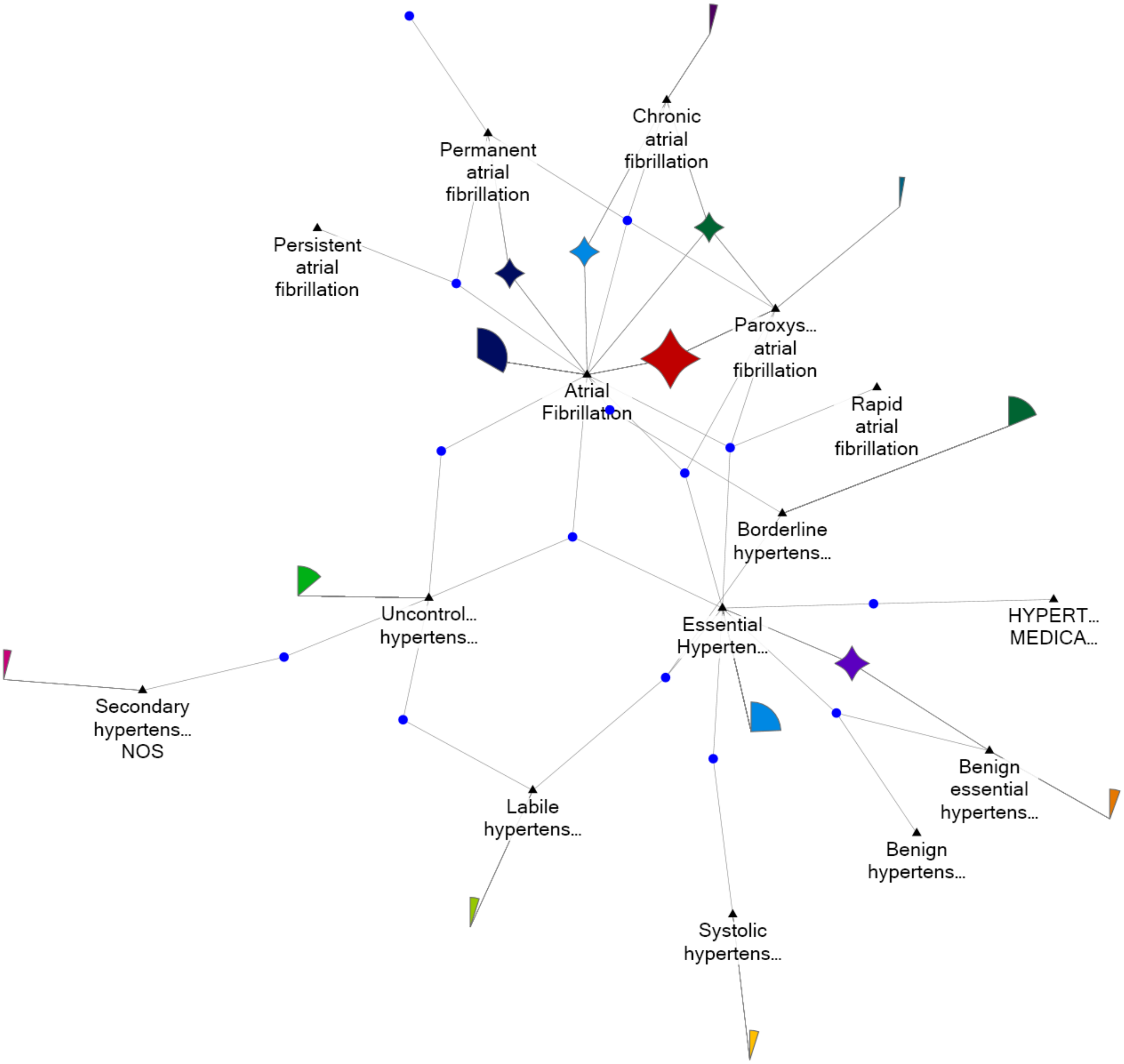


Fig. 12: Subnetwork of concepts matching “atrial fibrillation” or “hypertension” and their associated patients.

the top-right together connected to the same groups of patients. Our analysis indicates that doctors use these terms interchangeably. Another interesting finding is that the atrial fibrillation terms are not as directly related to the hypertension terms as Dr. Powsner had expected. There is a lack of connector motifs between these terms, and only four patients connect them.

3) Thrombosis and Stroke: Next, we looked at the subset of concepts that contained "thromb" or "stroke" and used the Select Subgraphs feature to find all the neighboring concepts via their intermediary patients. We colored the concepts matching "thromb" lime green and the ones matching "stroke" orange, while the neighboring concepts are black. This visualization is shown in Fig. 13. There were some cute findings from the visualization, including that the anticoagulant drugs grouped together geometrically in the bottom-left: "warfarin", "lovenox", "heparin", and "coumadin". Similarly, the platelet inhibitors "asprin" and "plavix" grouped together in the center-left. Many patients unsurprisingly shared the "asprin" and various thrombosis concepts.

There were some more disturbing revelations though. Many patients had deep vein thrombosis concepts but were not associated with any treatments. 17 of these are shown in the purple fan motif on the far right, and another 14 are visible in the dark green 3-connector motif nearby. Similarly, there were two untreated patients with ischemic strokes and another two with multiple strokes. It is not clear from these records whether all these disconcerting patients had no treatment for their conditions, whether it was simply a historical condition on their chart, or whether there was some sort of natural language processing failure when extracting their records.

IV. GROUP-IN-A-BOX LAYOUTS

A. Technique

In contrast to motif simplification, in which functionally equivalent nodes and edges are replaced by representative glyphs, we have also explored the use of **meta-layouts** that highlight more general topology- or attribute-based groupings of the network. These groups can be difficult to understand using the standard tools of color, shape, or convex hulls when faced with dense, intermingled topologic clusters. In this visualizations, it can be difficult to understand the size of each group, its internal structure, and its ties to other groups. Instead, we use **Group-in-a-Box layouts** [Cha+13; SD12; Rod+11; Dun13] that display groups individually to more clearly show membership, topology, and inter-group relationships.

In NodeXL we have several variants to more clearly show group relationships, each best suited to a range of topologies. The **Treemap Group-in-a-Box Layout** segments groups using a **Treemap**, which is space-filling but often separates related groups, drawing long edges which overlap other groups unnecessarily. This is visible in Fig. 14b as the crossing and overlapping meta-edges that represent the combined inter-group edges. The **Croissant Group-in-a-Box layout**, shown in Fig. 14c, puts the largest group at the top and wraps the remainder around three sides based on their connectivity. This effectively displays large groups, though more smaller groups are better shown using the **Donut Group-in-a-Box layout** (not shown here) which places the largest group in the center and arranges others around the perimeter. Finally, the **Force-Directed Group-in-a-Box layout** (Fig. 14d) arranges groups based on their aggregate ties and eliminates any overlap of their boxes.

Several case studies and experiments demonstrated [Cha+13; Dun13] that Group-in-a-Box layouts more clearly show (1) topology within groups, (2) group membership and size, and (3) aggregate relationships between groups. Group-in-a-Box layouts are particularly effective for large networks, where high density and finite screen space limit effective network visualizations.

B. Exploration

We also applied the three Group-in-a-Box meta-layouts to the network of patients and concepts from their discharge reports, originally discussed in Section III. After applying the Clauset-Newman-Moore topologic clustering algorithm [CNM04] to the network from Fig. 3, the standard color-coding approach produced the visualization shown in Fig. 15. The many densely connected clusters here are difficult to interpret. Note that standard clustering algorithms may not be as effective for analyzing networks with multiple node types, like this one of patients and concepts.

The Group-in-a-Box layouts, on the other hand, nicely segment these clusters. First, the Treemap GIB layout shown in Fig. 16 enables us to see the internal structure of each cluster. We have large clusters around our two egos in the network, the concepts "hops5325" and "orch7323" shown in orange. There is another large cluster to the top-right, as well as several smaller ones. Each of these clusters consist of several patients and a range of concepts associated with them. However, the Treemap layout prevents us from seeing the ties between clusters easily.

The Croissant-Donut layout, in this case choosing the Croissant variant, is shown in Fig. 17. This layout does somewhat better at removing the overlap of the meta-edges between groups though has worse aspect ratios for the group boxes. The pure Force-Directed approach, shown in Fig. 18, does even better at showing the group ties and maintains square group boxes, though group internal structure is a bit less discernable than in the Treemap layout.

One interesting combination is to use one of the Group-in-a-Box layouts with the motif simplification techniques we presented in Section III. We combined the node positions given by the Force-Directed GIB layout with the simplified motif glyphs, resulting in the visualization in Fig. 19. Due to technical limitations in the implementation, these approaches are not completely complimentary. For example, the edges between groups are shown and the group boxes have disappeared. However, the group and node positions are maintained. We can see which groups have large fan and connector motifs of similar concepts and could drill into them on a per-group basis. Future development, especially the inclusion of hierarchical or nested groups in NodeXL, could enable more effective combinations of these approaches.

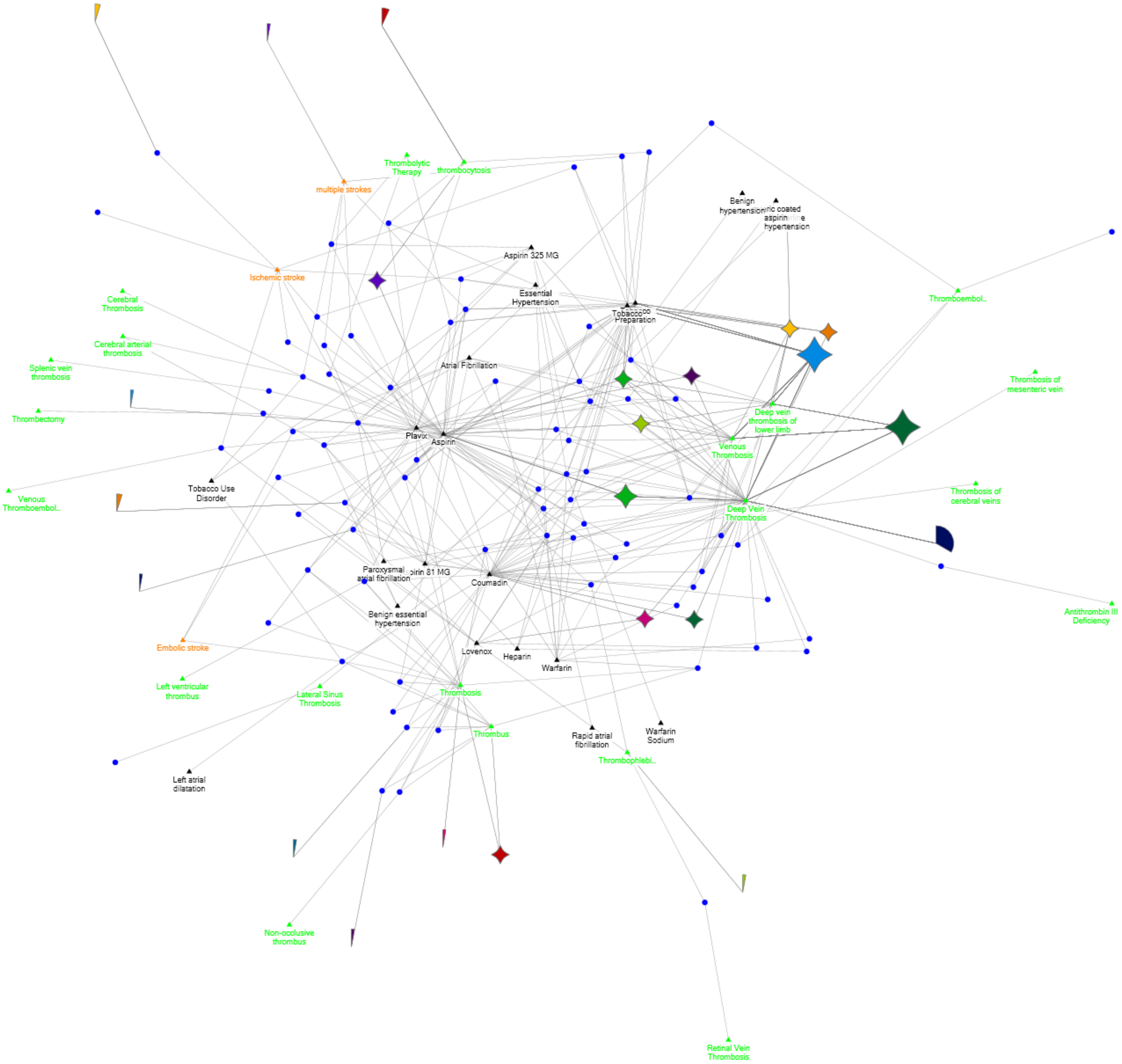
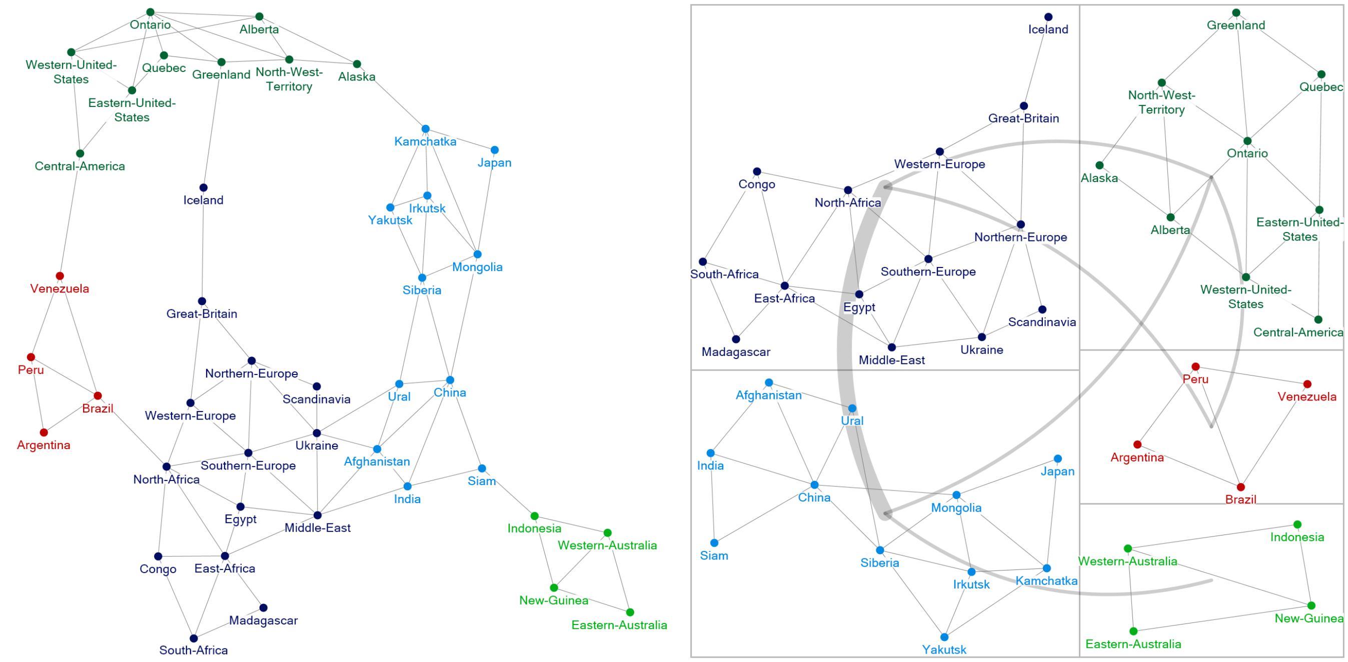
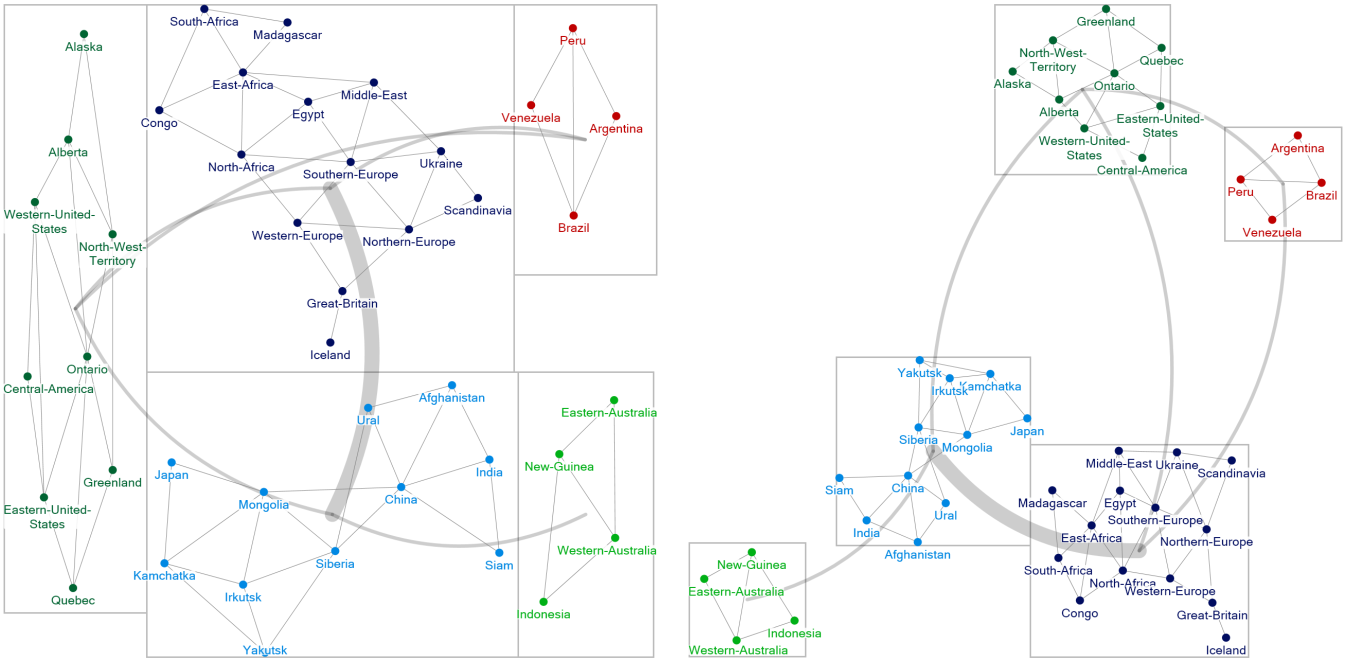


Fig. 13: Subnetwork of concepts matching “thromb” or “stroke”, their associated patients, and their neighboring concepts via intermediary patients.



(a) Standard node-link visualization (b) Treemap GIB layout



(c) Croissant-Donut GIB layout (d) Force-Directed GIB layout

Fig. 14: The network for the board game Risk, where nodes are countries and edges indicate legal movements. Nodes are laid out using Harel-Koren FMS [HK02], clustered and colored using the Clauset-Newman-Moore topologic clustering algorithm [CNM04]. Inter-group edges are combined into thick meta-edges. (a) shows the initial visualization, while the others show the three Group-in-a-Box (GIB) layout variants. See [Dun13] for more details and analysis.

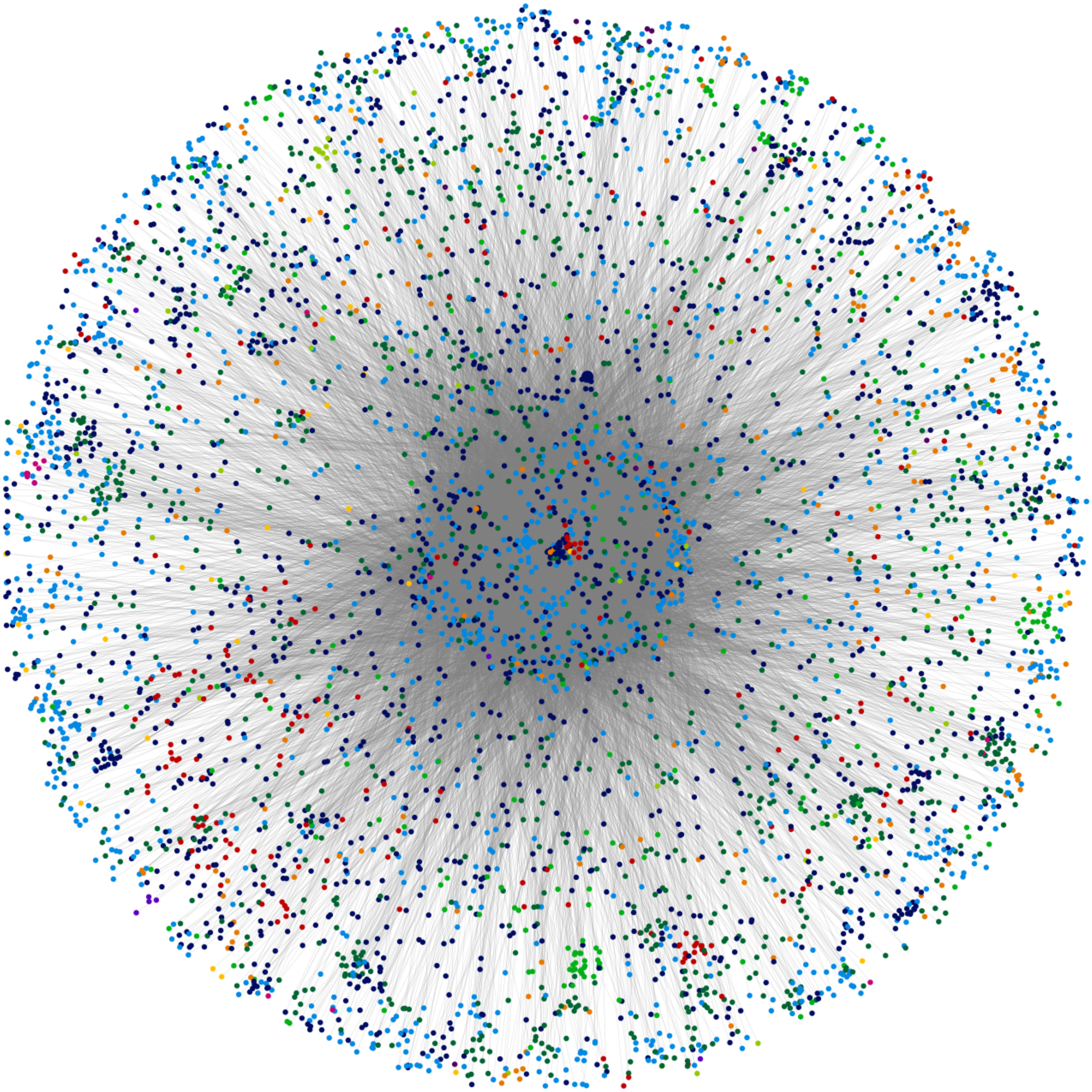


Fig. 15: Patients and concepts related to the “hops5325” and “orch7323” medications from Fig. 3. Nodes are grouped using the Clauset-Newman-Moore topologic clustering algorithm [CNM04] and colored accordingly.

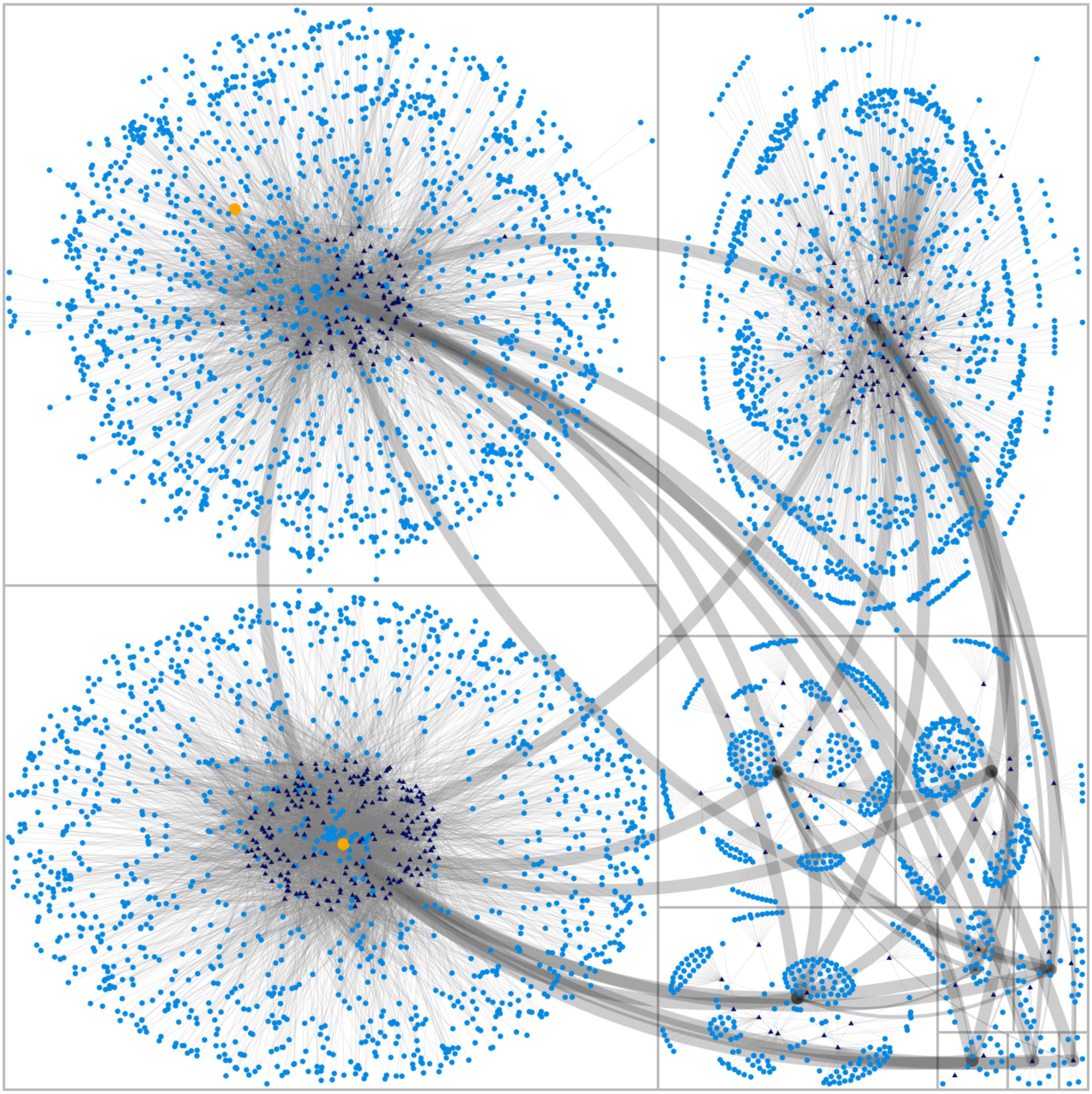


Fig. 16: Patients, concepts, and clusters from Fig. 15, shown in the Treemap Group-in-a-Box layout. Our ego concepts, “hops5325” and “orch7323”, are shown in orange in the largest clusters.

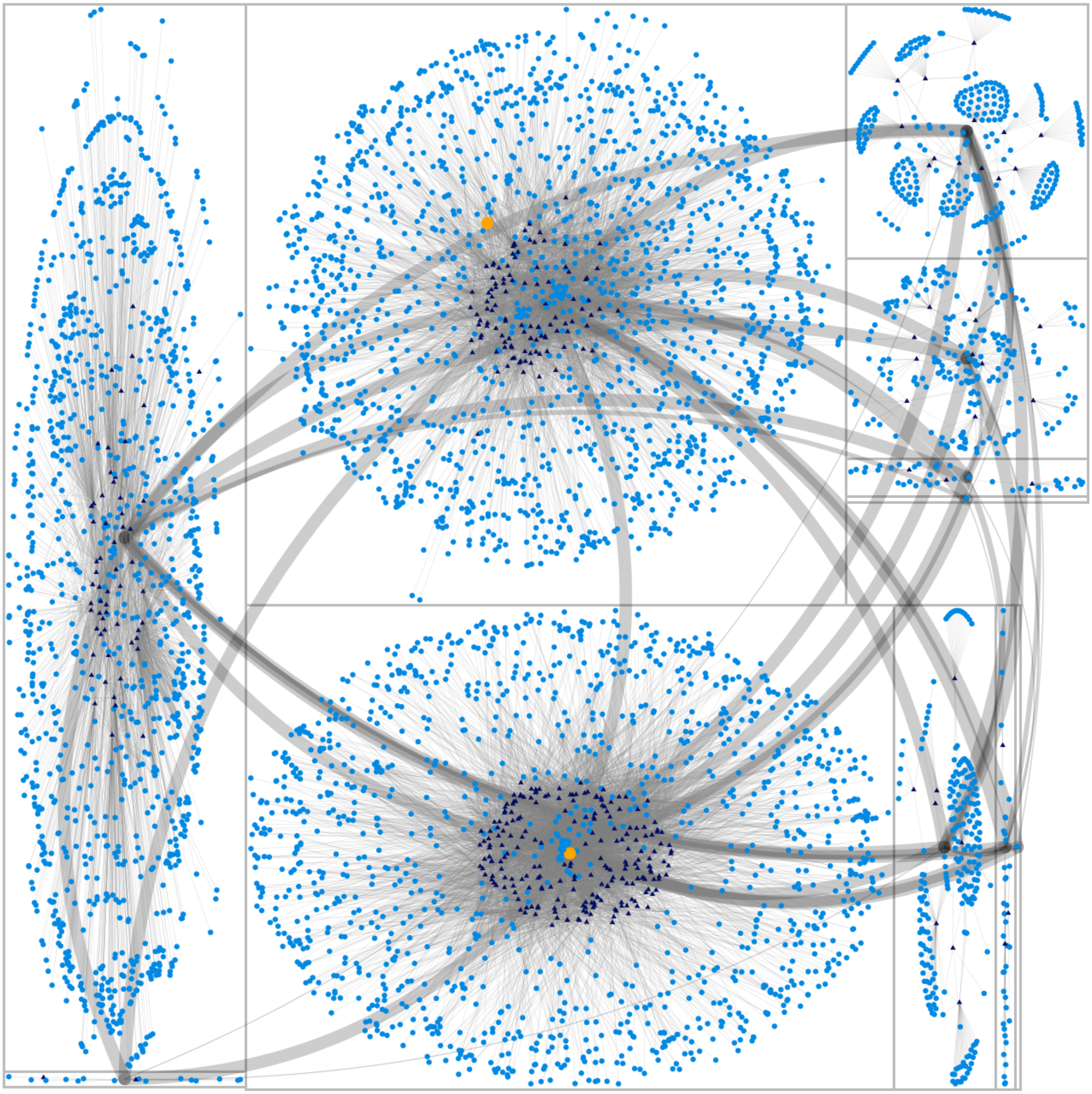


Fig. 17: Patients, concepts, and clusters from Fig. 15, shown in the Croissant-Donut Group-in-a-Box layout. In this case the Croissant variant was chosen automatically. Our ego concepts, “hops5325” and “orch7323”, are shown in orange in the largest clusters.

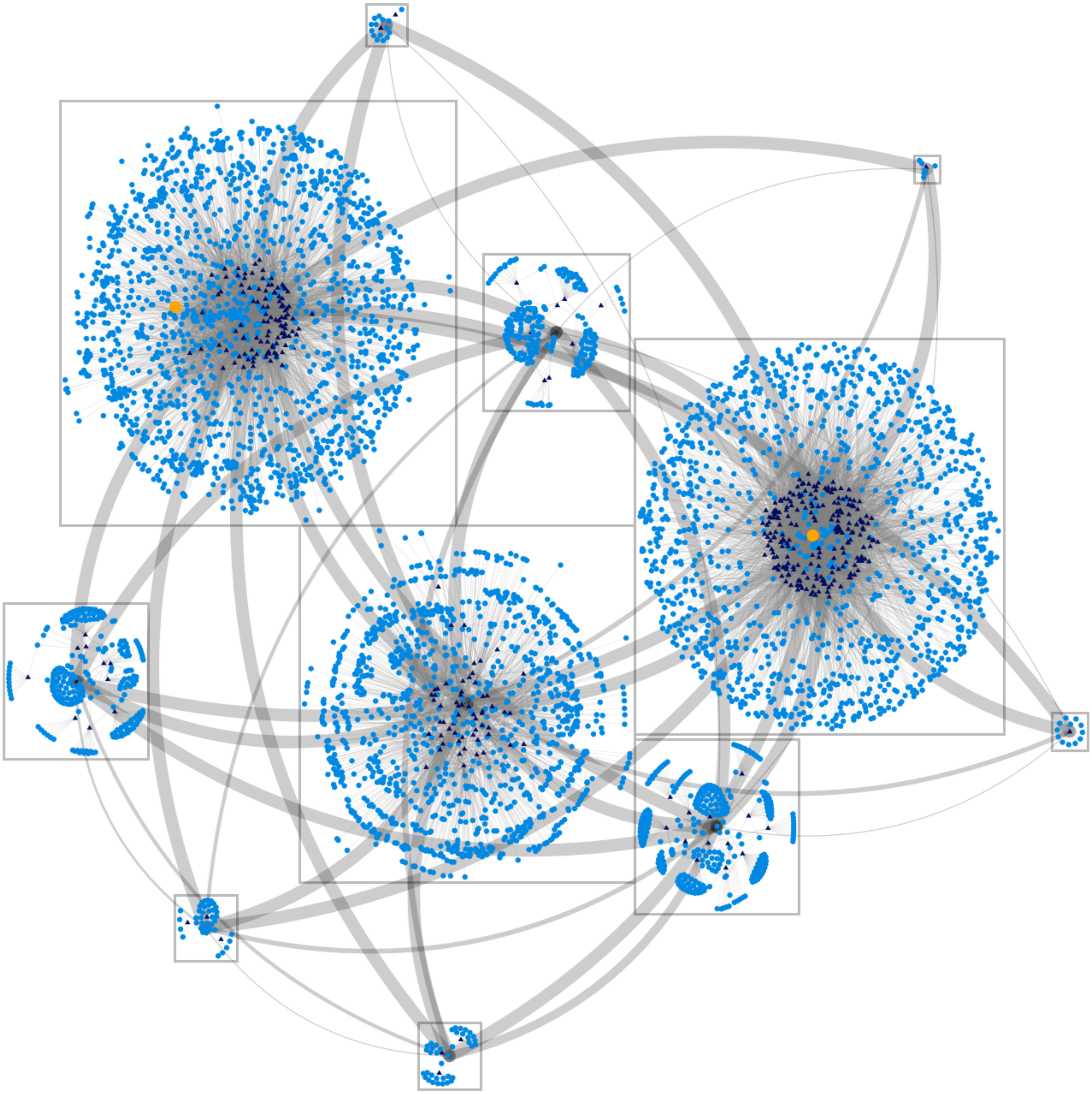


Fig. 18: Patients, concepts, and clusters from Fig. 15, shown in the Force-Directed Group-in-a-Box layout. Our ego concepts, “hops5325” and “orch7323”, are shown in orange in the largest clusters. The initial space-filling factor is 50%.

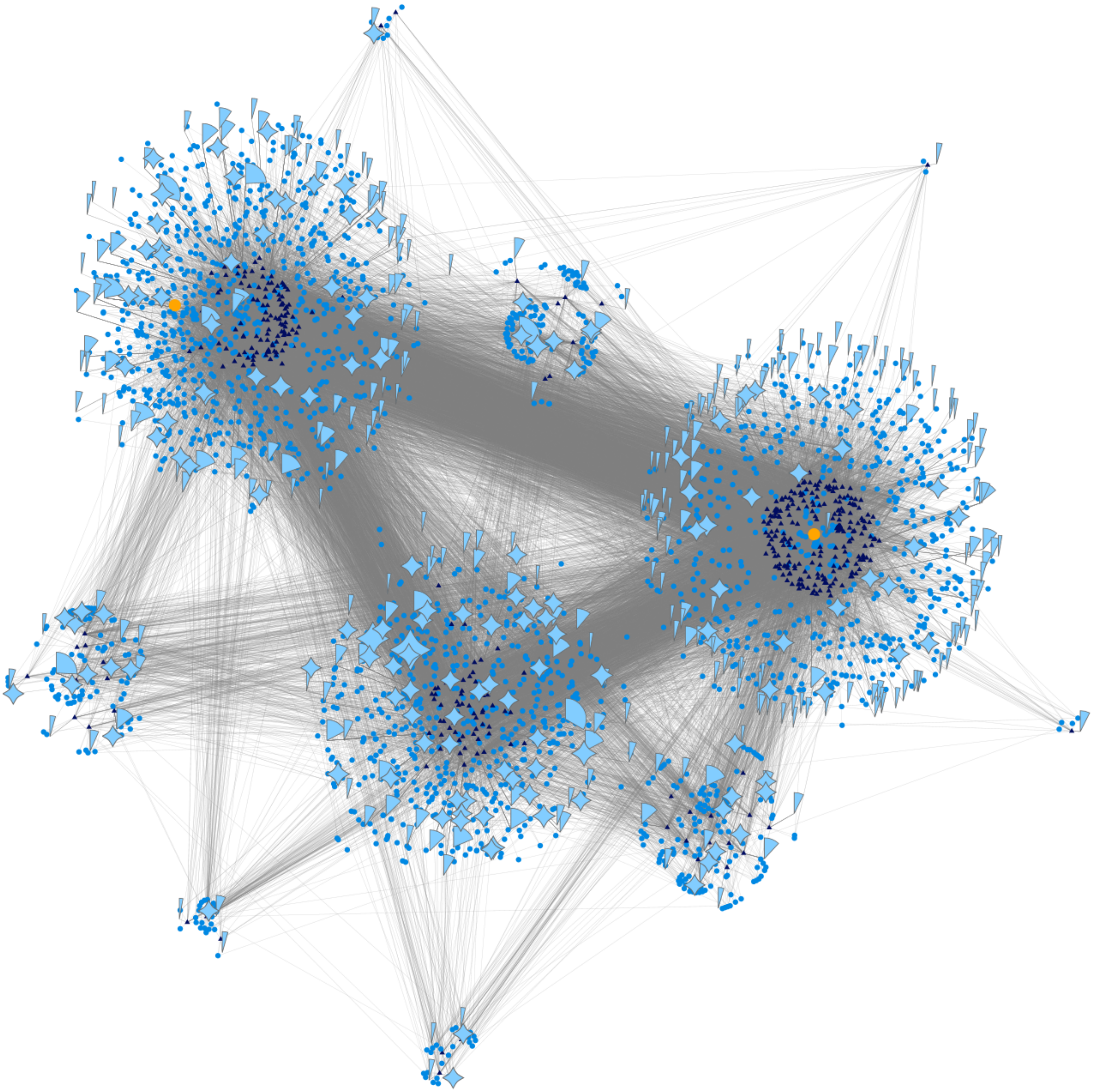


Fig. 19: Patients, concepts, and clusters from Fig. 15, shown in the Force-Directed Group-in-a-Box layout but without the group boxes. The underlying edges are visible. The motif simplification technique is applied as well.

References

[Cha+13] Snigdha Chaturvedi, Zahra Ashktorab, Cody Dunne, Rajan Zacharia, and Ben Shneiderman. *Group-in-a-Box layouts for visualizing network communities and their ties*. Under submission. 2013 (cit. on p. 14).

[CNM04] Aaron Clauset, Mark E. J. Newman, and Cristopher Moore. *Finding community structure in very large networks*. In: *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics* 70 (6 2004), p. 066111. DOI: 10.1103/PhysRevE.70.066111 (cit. on pp. 14, 16, 17).

[DS13] Cody Dunne and Ben Shneiderman. *Motif simplification: improving network visualization readability with fan, connector, and clique glyphs*. In: *CHI '13: Proc. SIGCHI Conference on Human Factors in Computing Systems*. CHI '13. 2013, pp. 3247–3256. DOI: 10.1145/2470654.2466444 (cit. on p. 2).

[Dun13] Cody Dunne. *Measuring and improving the readability of network visualizations*. PhD thesis. University of Maryland, 2013 (cit. on pp. 2, 14, 16).

[HK02] David Harel and Yehuda Koren. *A fast multiscale method for drawing large graphs*. In: *JGAA: Journal of Graph Algorithms and Applications* 6.3 (2002), pp. 179–202. DOI: 10.7155/jgaa.00051 (cit. on pp. 2, 3, 16).

[HSS11] Derek Hansen, Ben Shneiderman, and Mark Smith. *Analyzing social media networks with NodeXL: Insights from a connected world*. Ed. by Mary James and David Bevans. Morgan Kaufmann, 2011 (cit. on p. 1).

[Rod+11] Eduarda Mendes Rodrigues, Natasa Milic-Frayling, Marc Smith, Ben Shneiderman, and Derek Hansen. *Group-in-a-Box layout for multi-faceted analysis of communities*. In: *SocialCom '11: Proc. 2011 IEEE 3rd International Conference on Social Computing*. 2011, pp. 354–361. DOI: 10.1109/PASSAT/SocialCom.2011.139 (cit. on p. 14).

[SD12] Ben Shneiderman and Cody Dunne. *Interactive network exploration to derive insights: Filtering, clustering, grouping, and simplification*. In: *GD '12: Proc. 20th International Symposium on Graph Drawing*. Vol. 7704. Keynote. 2012, pp. 2–18. DOI: 10.1007/978-3-642-36763-2_2 (cit. on pp. 2, 14).

[Smi+09] Marc Smith, Ben Shneiderman, Natasa Milic-Frayling, Eduarda Mendes Rodrigues, Vladimir Barash, Cody Dunne, Tony Capone, Adam Perer, and Eric Gleave. *Analyzing (social media) networks with NodeXL*. In: *C&T '09: Proc. 4th International Conference on Communities and Technologies*. 2009, pp. 255–264. DOI: 10.1145/1556460.1556497 (cit. on p. 1).

[Smi+10] Marc Smith, Ben Shneiderman, Natasa Milic-Frayling, Eduarda M. Rodrigues, Vladimir Barash, Cody Dunne, Tony Capone, Adam Perer, and Eric Gleave. *NodeXL: A free and open network overview, discovery and exploration add-in for Excel 2007/2010*. Social Media Research Foundation. 2010. URL: http://nodexl.codeplex.com (cit. on pp. 1, 2).

[Smi+13] Marc Smith, Ben Shneiderman, Natasa Milic-Frayling, Eduarda M. Rodrigues, Vladimir Barash, Cody Dunne, Tony Capone, Adam Perer, and Eric Gleave. *NodeXLGraph Gallery*. Social Media Research Foundation. 2013. URL: http://nodexlgraphgallery.org/ (cit. on p. 1).