DYNAMIC, MULTI-LEVEL NETWORK MODELS OF CLINICAL TRIALS

MARCO D. SORANI[1], GEOFFREY T. MANLEY[1],
J. CLAUDE HEMPHILL[1] AND SERGIO E. BARANZINI[2]
[1]Department of Neurological Surgery and [2]Department of Neurology
University of California, San Francisco
San Francisco, CA 94110, USA
Emails: soranim@pharmacy.ucsf.edu, manleyg@neurosurg.ucsf.edu,
chemphill@sfgh.ucsf.edu, sebaran@cgl.ucsf.edu

While networks models have often been applied to complex biological systems, they are increasingly being implemented to investigate clinical questions. Clinical trials have been studied extensively by traditional statistical methods but never, to our knowledge, using networks. We obtained data for 6,847 clinical trials from five "Nervous System Diseases" (NSD) and five "Behaviors and Mental Disorders" (BMD) from the clinicaltrials.gov registry. We constructed networks of diseases and interventions for visualization and analysis using Cytoscape software. To standardize nomenclature and enable multi-level annotation, we used MeSH and UMLS terms. We then constructed separate BMD and NSD networks to study dynamics over time. To assess how topology features related to clinical significance, we constructed a sub-network of Multiple Sclerosis and Alzheimer's trials and identified which trials had been published in high-profile medical journals. We found that the BMD network has evolved into a large, decentralized topology and does not distinctly reflect the five diseases by which it was defined, while the NSD network does, though other diseases and sub-phenotypes have emerged as areas of research. We also found that high-profile trials have distinctive network characteristics. Future work is needed to address mathematical questions such as scale-dependence of network features, clinical questions such as trial design optimization, and methodological questions such as data quality improvement.

1. Background

Network models can reveal complex relationships in large data sets, and network topologies have been shown to share remarkably consistent features across diverse fields of study [1-3]. These features include scale free and small world properties, preferential attachment growth dynamics, vulnerability to perturbations of network hubs, and modularity. Network approaches are particularly revealing when they can describe systems in their entirety, incorporate quantitative and computable measurements, integrate multiple levels of detail, and capture the dynamics of a system over time [4]. Network models have often been applied to biological systems. For example, studies of gene-disease associations have interrogated genetic similarities among auto-immune diseases [5] and relationships between metabolic diseases and co-morbitities [6]. Increasingly, networks are also being implemented to investigate clinical and social questions. Recent studies have investigated the dynamics of infectious disease transmission [7], workflow in the intensive care unit [8], and collaborations resulting in publication [9,10].

Clinical trials, the gold standard of clinical research, have long been studied by traditional statistical methods [11-13] but never, to our knowledge, using network models. This is ironic since clinical collaborations have long been operated and even referred to as "networks" [14,15]. However, with requirements by journal editors [16] and U.S. Federal law [17] that certain trials be registered, and with the growth of public-access

registries, it is now possible to apply network models to clinical trial data. Furthermore, given the central role of clinical trials in translational research, network modeling may provide useful insights to address challenges such as designing comprehensive but non-redundant research programs and building on existing knowledge to design new trials.

In this study, we apply network models to characterize the dynamics and multi-level structure of a large set of data from clinical trials in nervous system diseases and behavioral and mental disorders obtained from the clinicaltrials.gov registry (http://www.clinicaltrials.gov). We also address challenges related to defining nomenclature and scope of clinical trial networks. We hypothesize that, as in other types of networks, hubs are functionally important, and networks grow according to a preferential attachment model. To test this, we describe features of these multi-level networks, including topological parameters, characteristics of network hubs and clusters, and the dynamics of clinical trial networks over time. We find that different disease types demonstrate divergent network topologies over time, and we observe distinct characteristics of clinically "influential" trials. These findings will be useful to assess areas of emphasis, overlap and omission in clinical research and funding programs as well as to identify relationships within and among disease phenotypes and therapeutic strategies.

2. The scope and construction of a clinical trials network

As the number of nodes in a network grows, the number of edges among them can increase exponentially, making analyses computationally intensive. There were 91,813 trials in clinicaltrials.gov categorized into 22 categories of diseases and conditions, as of 6/24/2010, and in the category of "Nervous System Diseases" alone, there were 65,462 non-unique trials in 506 specific diseases and conditions (Figure 1). To define a computationally tractable and clinically interpretable system, we downloaded 6,847 non-unique trials from five conditions categorized as "Nervous System Diseases" (NSD) and five conditions categorized as "Behaviors and Mental Disorders" (BMD). The NSD conditions were "Alzheimer Disease" (AD) (687 studies), "Brain Injuries" (505), "Multiple Sclerosis" (MS) (536), "Parkinson Disease" (PD) (629), and "Stroke" (1070); the BMD conditions were "Alcoholism" (357), "Attention Deficit and Disruptive Behavior Disorders"(444), "Bipolar Disorder" (584), "Schizophrenia" (1271), and "Smoking" (764). The NSD and BMD categories had similar total numbers of trials (3427 v. 3420) for later comparisons.

For each trial, we downloaded multiple parameters and performed several types of re-coding on the data (Table 1). We obtained National Clinical Trial (NCT) ID, Recruitment Status, Condition, Intervention, Sponsor, Study Type, Start Date and Completion Date for each trial. Intervention and Sponsor data frequently contained Unicode characters which we converted to ASCII text. Intervention attributes were nested (e.g., Drug: Aspirin) and were parsed to derive an Intervention Type field (e.g., Drug). Multiple fields contained more than one attribute per row in a pipe ("|") delimited manner (e.g., Drug: Aspirin|Drug: Codeine) and were also parsed. Other parameters such as study design and outcome are available, but they are highly heterogeneous and not readily computable.

Fig. 1

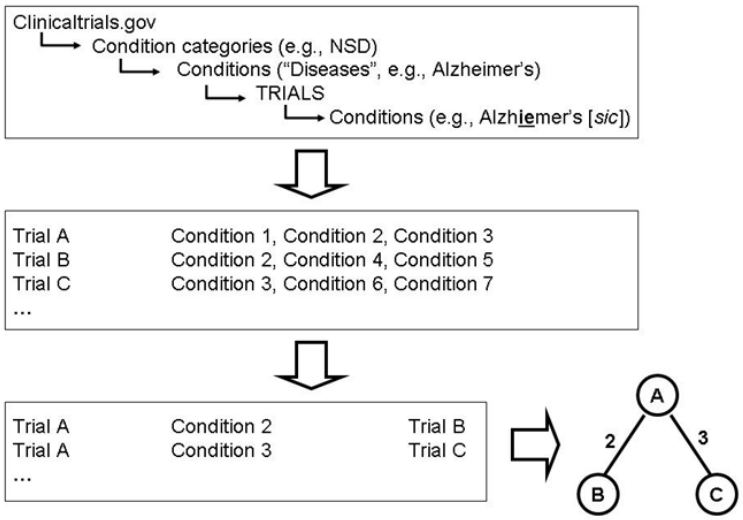

Fig. 1. Schematic representation of clinical trial data converted from tabular to graph structure. Clinicaltrials.gov contains disease data in a nested, partly ambiguous hierarchy. Trials are organized into categories such as Nervous System Diseases (NSD). These categories include standardized Conditions which include specific trials. Specific trials can include multiple free-form Conditions entered by investigators--please note the hypothetical typographical error. Since the term "Conditions" is used twice, we subsequently refer to the higher Conditions as "Diseases". From the tabular layout, files of node-edge-node triplets were generated for import into Cytoscape.

Where possible, we standardized conditions to Medical Subject Headings (MeSH) (http://www.ncbi.nlm.nih.gov/mesh) and Unified Medical Language System (UMLS) (http://www.nlm.nih.gov/research/umls) nomenclature. Clinicaltrials.gov uses the term "Condition" for both their pre-defined diseases and the multiple free-form, mixed case text conditions submitted by investigators for each trial. Intervention data are also submitted by investigators as free-form text. For example, in multiple sclerosis, there are three pre-defined Diseases: "Multiple Sclerosis", "Multiple Sclerosis, Relapsing-Remitting", and "Multiple Sclerosis, Chronic Progressive". Diseases map to MeSH terms and identifiers such as "Multiple Sclerosis, Chronic Progressive" [C10.314.350.500.200]. By contrast, there may be dozens of free-form text Conditions, some of which reflect true sub-phenotypes [18] and some of which are cases of inconsistent nomenclature, as shown in the examples related to brain injuries in Table 2. To standardize free-form text to UMLS Concept Unique Identifiers (CUI), we used the Batch SemRep web tool (http://skr.nlm.nih.gov).

We then constructed networks for visualization and analysis using Cytoscape software (http://www.cytoscape.org). First, we implemented custom Python scripts (http://www.python.org) to convert tabular data into undirected graphs defined by node-edge-node triplets (e.g., Trial-Condition-Trial) (Figure 1). Briefly, we iterated through all Conditions in all trials to identify trials that studied Conditions in common. This same algorithm was applied to Interventions and Sponsors and could be applied to any trial parameter. We also generated node attribute files defined by "node = <type>" statements (e.g., "Trial1 = NSD"). Once we constructed the networks, we analyzed their topologies

visually and quantitatively. In this study, networks are displayed using the yFiles (http://www.yworks.com) Organic layout. Using the Network Analyzer plug-in [19] and custom code, we computed common topological parameters including node degree (i.e., the number of edges incident to the node) and others. We also identified hubs and clusters in the networks.

Table 1. Example record layout for a single trial. The Category, Disease and Intervention Type fields were derived from the clinicaltrials.gov primary data.

Parameter	Example value
NCT ID	NCT00167323
"Category"	BMD
"Disease"	Alcoholism
Recruitment Status	Completed
Intervention	Behavioral: Adherence therapy
"Intervention Type"	Behavioral
Study Type	Interventional
Start Date	Jul-03
Completed Date	Jul-07
Condition	Bipolar Disorder|Alcohol Use Disorder
Sponsors	University of Pittsburgh|National Institute on Alcohol Abuse and Alcoholism (NIAAA)

Table 2. Examples of non-standard nomenclature in the Condition field of trials in the "Brain Injuries" category, in order of frequency. UMLS identifiers and terms were assigned using the Batch SemRep web tool.

Standard UMLS term	UMLS CUI	Total	Free-text Condition	Subtotal
Traumatic Brain Injury	C0876926	173	Traumatic Brain Injury	148
			Mild Traumatic Brain Injury	10
			Severe Traumatic Brain Injury	9
			TBI (Traumatic Brain Injury)	6
Cerebral Palsy	C0007789	96	Cerebral Palsy	96
Brain Injuries	C0270611	82	Brain Injury	41
			Brain Injuries	33
			Brain Injuries Traumatic	5
			Acquired Brain Injury	3
Craniocerebral Trauma	C0018674	13	Craniocerebral Trauma	7
			Head Injury	6
Hypoxic-Ischemic Encephalopathy	C0752304	12	Hypoxic Ischemic Encephalopathy	7
			Hypoxic-Ischemic Encephalopathy	5
Subarachnoid Hemorrhage	C0038525	11	Subarachnoid Hemorrhage	11

3. Annotation of multiple network levels

A fundamental challenge in network modeling and systems biology is the integration of multiple "levels" of data, that is, data with different levels of granularity. We constructed multi-level networks to analyze trial Conditions, Interventions, and Sponsors. We defined four Condition levels, from the top down, as (1) the two clinicaltrials.gov categories (BMD and NSD) or "Both", (2) the 10 clinicaltrials.gov diseases or "Multiple", (3) the standardized UMLS CUIs, and (4) the free-form text Conditions. We defined three Intervention levels as (1) the Intervention type (e.g., "Drug" for the Intervention "Drug: Aspirin"), (2) the standardized CUIs of the Intervention, and (3) the free-form text Intervention. We defined two Sponsor levels as (1) sets of Sponsors and (2) individual Sponsors.

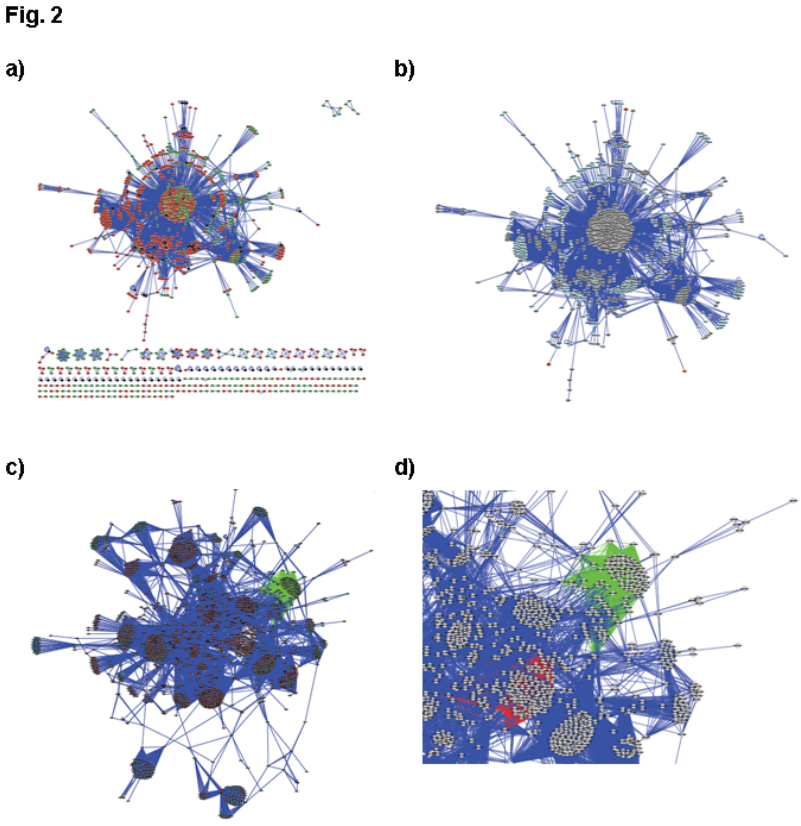

Fig. 2. Multi-level networks. Intervention networks are colored (a) by disease category, Behaviors and Mental Disorders (BMD) in red, Nervous System Diseases (NSD) in green, Both in black and (b) by intervention type ("Multiple" in gray, "Drug" in cyan). The central cluster represents placebo-controlled trials. There are more BMD trials represented than NSD trials and relatively few "Both" trials. Sponsor networks are colored (c) by category (nodes: BMD in red, NSD in green, Both in black) and (d) in an enlarged view of the upper right quadrant of (c), by the specific sponsor (edges: the National Institute of Mental Health in red, the National Institute of Neurological Disorders and Stroke in green).

To construct multi-level networks of Conditions, Interventions and Sponsors, we selected a subset of 2412 completed interventional trials (Figure 2). The Conditions network had 2202 nodes, similar to the Interventions and Sponsors networks (Table 3) but had 231,813 edges, a 4-5 fold increase compared to the other networks. A clear result of multi-level visualization is that, especially in large networks, higher levels of aggregation (e.g., disease category v. individual disease) are easier to interpret.

All networks had a large, primary connected component subdivided into more or less distinct sub-graphs which correspond to higher levels of aggregation, such as disease category. Clear clusters were also visible and correspond to lower levels of aggregation, such as placebo intervention or large sponsors.

Table 3. Similar topology parameters of completed interventional trial networks of Interventions and Sponsors. The clustering coefficient is a measure of the extent to which nodes in a graph cluster together.

Network	Nodes	Edges	Clustering coefficient	Diameter
Interventions	1372	61,749	0.74	9
Sponsors	2124	53,248	0.87	9

4. Assessment of network dynamics

Clinical trials provide an opportunity to study network dynamics because rich longitudinal data is available, and because multiple behaviors can be observed including new node and link formation as well as "death" (e.g., when a trial is withdrawn or terminated). We identified all trials in the registry that started after 1980 or completed before 2009. We did not include trials started after 2005 to control for possible lags in registration. We then constructed separate networks for BMD and NSD trials to compare their evolution over time, visually (Figure 3) and by topology metrics.

Trial network topologies diverged over time between BMD and NSD. Of note, for both the BMD and NSD networks, approximately 90% of trials in the registry began after 2000. Before 1995, there was a core BMD component and a separate co-morbidity component. Between 1996-2000, the schizophrenia\bipolar disorder core component grew, and alcohol and smoking trial clusters began to form, while attention deficit trials were a separate component. Since 2000, the BMD network has become large and decentralized and does not distinctly reflect the five major diseases by which it was defined. Before 1995, there was a greater total number of trials in NSD than in BMD, including a large stroke trial cluster, a more sparse cognitive disorders cluster, and a separate MS component. Between 1996-2000, the network clusters grew and became more integrated. Since 2000, the NSD network's clusters reflect the five major diseases by which it was defined. In addition, new disease trial clusters such as cerebral palsy have emerged as major areas of research, other areas like MS and Parkinson's Disease have spawned sub-phenotypes such as Secondary Progressive MS and Idiopathic PD, and more rare conditions have emerged. Notably, around 2005, first in BMD and then in NSD, the increase in completing trials began to exceed the number of starting trials in the registry.

Fig. 3

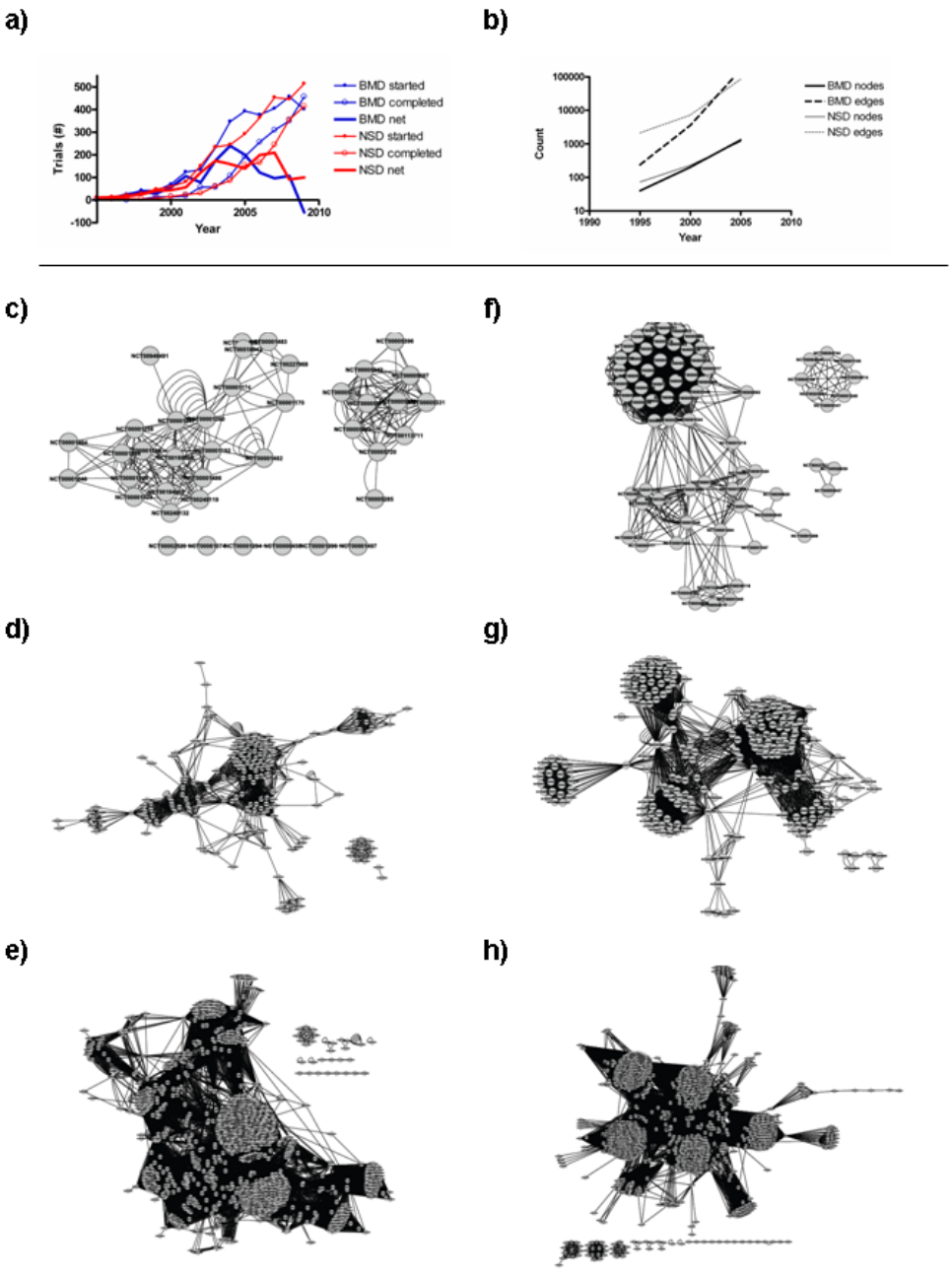

Fig. 3. Clinical trial network dynamics. (a) The number of trials started and completed in Behaviors and Mental Disorders (BMD) and Nervous System Diseases (NSD) rose slowly until about 2000 and then increased rapidly. (b) Sizes of both networks grew logarithmically, but the number of edges grew at different rates. Networks of BMD trials started (c) up until 1995, (d) between 1996-2000, and (e) between 2001-2005 are shown. Networks of NSD trials started (f) up until 1995, (g) between 1996-2000, and (h) between 2001-2005 are also shown.

These differences were reflected in network topology parameters. The BMD network's clustering coefficient increased over time (0.67, 0.85, 0.92) but was unchanged in the NSD network (0.91, 0.9, 0.93). In contrast, the BMD network's diameter and characteristic path length first rose and then fell, but both metrics rose steadily in the NSD network (4, 10, 6 v. 3, 5, 11 and 1.6, 3, 2.2 v. 1.8, 2.4, 2.5). Finally, the BMD network's density was unchanged over time (0.21, 0.16, 0.18) but fell in the NSD network (0.29, 0.17, 0.11).

5. Network characteristics of "influential" trials

We next examined how network topology features and dynamics related to clinical significance. First, we constructed a sub-network of the 479 MS and Alzheimer's trials in clinicaltrials.gov with a status of "Completed" and a reported start date. We then identified MS and Alzheimer's trials published in the Journal of the American Medical Association or the New England Journal of Medicine since 2005 to represent "influential" clinical trials. Figure 4 shows the topological characteristics of these influential studies relative to other trials in the context of Conditions, Interventions, and Sponsors networks.

Fig. 4

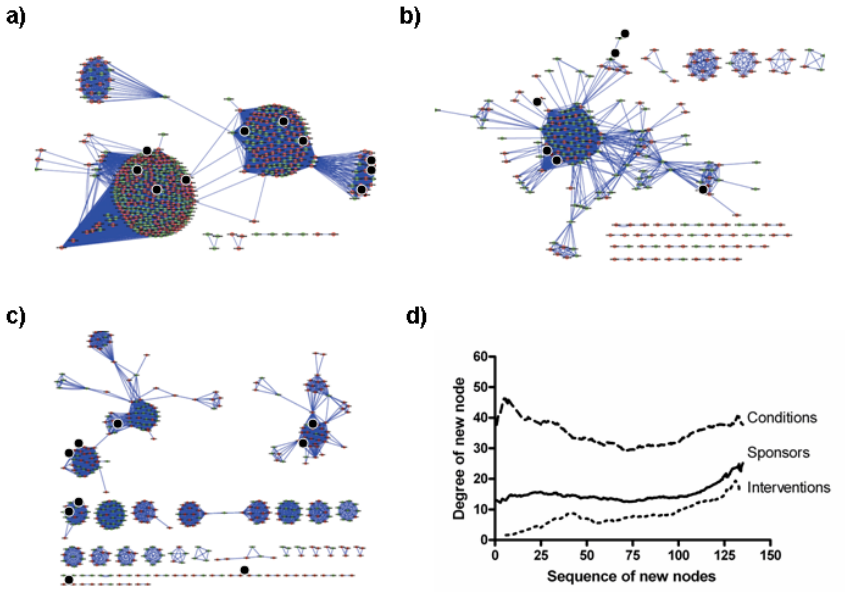

Fig. 4. Topological features of "influential" trials. (a) In the Conditions network, influential trials (black) are members of large disease clusters (AD, MS, Relapsing Remitting MS) near both prior (red) and subsequent (green) trials. (b) In the Interventions network, influential trials are in the network periphery, representing novel approaches, and are primarily near subsequent trials, supporting their influence on future studies. These include four trials in RRMS using fingolimod, interferon beta-1a, cladribine, and rituximab. (c) In the Sponsors network, influential trials are dispersed across various clusters of trials with common sponsors, which represent academic and industrial collaborations. They also occur primarily near prior trials, suggesting a culmination of previous efforts. (d) While the current degree of new nodes in the Intervention and Sponsor networks increased steadily as networks grew, the degree of new nodes in the Condition network started high and fell before rising again.

We set out to understand the dynamics of these three networks and, in particular, to describe when high-degree network hubs joined the networks. We first calculated degree distributions for the three networks in their current state. We then sorted the trials, that is the network nodes, by their start dates. Finally, we fit a coarse lowess spline to the three network series to identify trends in the data (Figure 4d). The degree of new nodes in the Intervention and Sponsor networks increased steadily, suggesting that early interventions were often abandoned and early sponsors often left the field, while later entrants formed more connections. The degree of new nodes in the Condition network started high, like a preferential attachment model, but fell before rising again, suggesting that medically relevant conditions were identified early and more have been defined recently.

We have defined influential trials using just one out of a multitude of possible ways. The increasing number of connected components in the networks (n=1, 6 and 16, for >3 nodes) supports the hypothesis that innovation may spread most easily among disease areas but may disseminate more slowly among groups studying different interventions, and information may flow with some difficulty across unconnected sponsor groups.

6. Discussion

In this study, we present the first network-based analysis, to our knowledge, of clinical trial data. Using a large set of data from clinical trials in nervous system diseases and behavioral and mental disorders obtained from the clinicaltrials.gov registry, we examine the topological parameters, network features, and longitudinal dynamics of clinical trial conditions, interventions and sponsors. We propose solutions to defining nomenclature and scope for constructing clinical trial networks. We hypothesized that, as in other types of networks, hubs and clusters are functionally important, and networks grow according to a preferential attachment model. We found that the role of network hubs was more similar among conditions and sponsors, since those hubs had functionally dominant roles, whereas, aside from the placebo cluster in the interventions network, the interpretation of functional importance was less clear. We also found that networks of different disease categories grew in divergent manners, and networks demonstrated variant models of preferential attachment.

Inconsistent data quality has previously been identified as an impediment to the construction of biological networks [20,21] and clinical databases [22,23]. One of the major challenges to studying human diseases computationally is the development of vocabularies and ontologies that realistically reflect the complex inter-relationships among phenotypes. Multiple solutions to this problem have been reported including mapping diseases to OMIM disorders [24], Medicare records [6], MeSH terms [25], ICD-9 codes and other ontologies. Clinicaltrials.gov implements MeSH terminology at the upper levels of its disease classification system but allows submission of free-form text at the lower levels. Other attributes such as interventions are unstructured but could be mapped to reference data sets such as RxNorm [26] or Drugs@FDA (http://www.accessdata.fda.gov). Still other attributes are submitted by trial sponsors with limited safeguards to ensure the accuracy and consistency of terminology.

We used clinicaltrials.gov and UMLS terms to standardize trial nomenclature and enable multi-level analysis. Multi-level analysis of trials may be useful when sponsors evaluate trials "upward" as components in an overall research program and "downward"

as collections of individual patients. In both cases, issues of membership overlap and hierarchy may be encountered. Ahn et al. [27] constructed communities that incorporate overlap and hierarchical organization in biological and social networks. We addressed overlap by implementing classes such as "Both" for disease categories and "Multiple" for diseases.

Understanding the dynamic "evolution" of clinical trials from a systems perspective is similar to a phylogenetic analysis of ecosystems and may be useful in understanding the emergence, persistence, diversification and modularity of clinical research, particularly given the "noise" we have described in this type of data [28]. Many models have been proposed to describe dynamics in different types of network systems. These models include preferential attachment and linear distance dependence in internet topology [29], duplication-mutation schemes in the E. coli genetic network [30], modified preferential attachment in sexual contact networks [31], asymmetric disassembly for contraction and preferential attachment for re-growth in the New York garment industry [32], and anti-preferential attachment in protein-protein interaction [33]. In the context of clinical trials, one might expect that a growth model similar to preferential attachment might hold true for networks of conditions and sponsors, where it may be the case that "the rich get richer". However, for networks of interventions, one might expect an altogether different growth model, where it is less likely that trials will be initiated for an existing intervention, since an intervention will eventually either fail and disappear or succeed and no longer require new studies, though in both cases it might be introduced into new indications.

Studying the evolution of clinical trial networks can provide insight into mechanisms of knowledge flow, just as studying the spatial-temporal transmission of infectious disease might provide insight into mechanisms of communicability. While the flow of a virtual entity like information through a trial network is different from the flow of a physical entity like electricity through a power grid, both require sources and connectivity. An example of a knowledge source is the release of information via Pubmed, clinicaltrials.gov, or perhaps a conference, highlighting the importance of registering and reporting results of trials, including negative studies, in a complete and timely manner. An example of knowledge connectivity was described in Figure 4 where the differing connections between components in the three networks may have implications on how knowledge from influential trials is disseminated. This is significant because the average time to take a new therapeutic compound from discovery to commercialization in the U.S. is nearly 13 years, up from less than eight years in the 1960s. Opening a trial requires a median of approximately 2.5 years simply to begin patient accruals, not to complete the trial [34].

The primary limitation of this study is the reliance on potentially ambiguous categorizations and nomenclature of study conditions. For example, there are 536 studies in the pre-defined "Multiple Sclerosis" category, 181 studies under "Multiple Sclerosis, Relapsing-Remitting", and 26 studies under "Multiple Sclerosis, Chronic Progressive". However, searching with the text term "Multiple Sclerosis" returns 585 studies, while searching for the terms independently returns 625 studies. Other recognized phenotypes such as Primary Progressive MS are not explicitly defined. Clearly, the definition of disease nomenclature is an ongoing effort.

There are several potential future directions for this work. We chose to take a disease-centric approach by focusing on relationships among trials in neurological diseases. Alternative approaches could include expanding scope, by looking at all diseases; altering scope, by looking at a different set of diseases or a specific class of interventions; or altering the network model, by focusing on relationships using some unit other than trials. There are many opportunities to study other aspects of network dynamics. For example, in studies of corporate networks, the merging and splitting of nodes can represent acquisitions and spin-offs. Similarly, merging and splitting of nodes in clinical trial networks could reflect evolving understanding of sub-phenotypes of disease or differences in drug mechanisms of action. To further integrate multiple levels of detail into network models of clinical trials, it would be extremely useful to have patient-level data, but this may be difficult to obtain since such data is typically reported at a summary level, if at all. Other issues to investigate include scale-dependence [35] and network vulnerability [36]. For example, in the same way that other types of networks such as power grids or the Internet might experience cascading failure, what might happen to the practice of medicine if the findings of a "hub" trial are called into question?

In conclusion, we have presented the first network-based analysis of public clinical trials data. We defined a large set of trials in neurological conditions using data from clinicaltrials.gov. We analyzed multi-level models that integrated levels of granularity of trial conditions, interventions, and sponsors. We also analyzed dynamic models of network evolution over time. In both cases, we performed visual and topological evaluations. We highlight opportunities to make trial nomenclature more consistent and computable, we describe divergent network topologies over time in different disease types, and we identify characteristics of clinically "influential" trials in neurology using network models.

Acknowledgments

We thank Leonard Apeltsin and Michael Keiser for helpful discussions.

References

1. A. Clauset, C. Moore, M. E. Newman, *Nature.* **453**, 7191 (2008).
2. C. T. Butts, *Science.* **325**, 5939 (2009).
3. A. Tero, S. Takagi, T. Saigusa, K. Ito, D. P. Bebber, M. D. Fricker, K. Yumiki, R. Kobayashi and T. Nakagaki, *Science.* **327**, 5964 (2010).
4. [podcast] S. Mirsky and L. Hood, *Sci. Am.* July 11, (2007).
5. S. E. Baranzini, *Curr. Opin. Immunol.* **21**, 6 (2009).
6. D. S. Lee, J. Park, K. A. Kay, N. A. Christakis, Z. N. Oltvai and A. L. Barabasi, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 29 (2008).
7. S. Riley, *Science.* **316**, 5829 (2007).
8. J. E. Gray, D. A. Davis, D. M. Pursley, J. E. Smallcomb, A. Geva and N. V. Chawla, *Pediatrics.* **125**, 6 (2010).
9. S. A. Greenberg, *BMJ.* **339** (2009).
10. C. M. Morel, S. J. Serruya, G. O. Penna and R. Guimaraes, *PLoS Negl. Trop. Dis.* **3**, 8 (2009).

11. J. F. Scoggins and D. L. Patrick, *Contemp. Clin. Trials.* **30**, 4 (2009).
12. X. Cao, K. B. Maloney and V. Brusic, *Immunome Res.* **4** (2008).
13. A. R. Brunoni, L. Tadini and F. Fregni, *PLoS One.* **5**, 3 (2010).
14. Asthma Clinical Research Network. J. M. Drazen, E. Israel, H. A. Boushey, V. M. Chinchilli, J. V. Fahy, J. E. Fish, S. C. Lazarus, R. F. Lemanske, R. J. Martin, S. P. Peters, C. Sorkness and S. J. Szefler, *N. Engl. J. Med.* **335**, 12 (1996).
15. MAC Study Group of the Canadian HIV Trials Network. S. D. Shafran, J. Deschenes, M. Miller, P. Phillips and E. Toma, *N. Engl. J. Med.* **330**, 6 (1994).
16. F. Berenbaum, *Best Pract. Res. Clin. Rheumatol.* **24**, 1 (2010).
17. T. Tse, R. J. Williams and D. A. Zarin, *Chest.* **136**, 1 (2009).
18. G. Miller, *Science.* **328**, 5976 (2010).
19. Y. Assenov, F. Ramírez, S. E. Schelhorn, T. Lengauer and M. Albrecht, *Bioinformatics.* **24**, 2 (2008).
20. J. Y. Chen, S. Mamidipalli and T. Huan, *BMC Genomics.* **10**, Suppl 1 (2009).
21. M. L. Nahm, C. F. Pieper and M. M. Cunningham, *PLoS One.* **3**, 8 (2008).
22. J. D. Horbar and K. A. Leahy, *Control. Clin. Trials.* **16**, 1 (1995).
23. A. Lux, S. Kropf, E. Kleinemeier, M. Jurgensen and U. Thyen; DSD Network Working Group, *BMC Public Health.* **9** (2009).
24. K. I. Goh, M. E. Cusick, D. Valle, B. Childs, M. Vidal and A. L. Barabasi, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 21 (2007).
25. F. Barrenas, S. Chavali, P. Holme, R. Mobini and M. Benson, *PLoS One.* **4**, 11 (2009).
26. F. Parrish, N. Do, O. Bouhaddou and P. Warnekar, *AMIA Annu. Symp. Proc.* **1057** (2006).
27. Y. Y. Ahn, J. P. Bagrow and S. Lehmann, *Nature.* **466**, 7307 (2010).
28. S. Erten, X. Li, G. Bebek, J. Li and M. Koyuturk, *BMC Bioinformatics.* **10** (2009).
29. S. H. Yook, H. Jeong and A. L. Barabasi, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 21 (2002).
30. M. Middendorf, E. Ziv, C. Adams, J. Hom, R. Koytcheff, C. Levovitz, G. Woods, L. Chen and C. Wiggins, *BMC Bioinformatics.* **5 (**2004).
31. B. F. de Blasio, A. Svensson and F. Liljeros, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 26 (2007).
32. S. Saavedra, F. Reed-Tsochas and B. Uzzi, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 43 (2008).
33. W. K. Kim and E. M. Marcotte, *PLoS Comput. Biol.* **4**, 11 (2008).
34. D. Dilts, *Nat. Med.* **16**, 6 (2010).
35. N. J. Gotelli, G. Graves and C. Rahbek, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 5030 (2010).
36. S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley and S. Havlin, *Nature.* **464**, 7291 (2010).