NeuroImage 183 (2018) 872–883

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

The landscape of *NeuroImage*-ing research

Jordan D. Dworkin[a,*], Russell T. Shinohara[a], Danielle S. Bassett[b,c,d,e]

[a] Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
[b] Department of Bioengineering, School of Engineering and Applied Sciences, University of Pennsylvania, Philadelphia, PA, USA
[c] Department of Physics & Astronomy, College of Arts and Sciences, University of Pennsylvania, Philadelphia, PA, USA
[d] Department of Electrical & Systems Engineering, School of Engineering and Applied Sciences, University of Pennsylvania, Philadelphia, PA, USA
[e] Department of Neurology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

ARTICLE INFO

Keywords:
Knowledge network
Graph theory
Neuroimaging
Network science
Science of science

ABSTRACT

As the field of neuroimaging grows, it can be difficult for scientists within the field to gain and maintain a detailed understanding of its ever-changing landscape. While collaboration and citation networks highlight important contributions within the field, the roles of and relations among specific areas of study can remain quite opaque. Here, we apply techniques from network science to map the landscape of neuroimaging research documented in the journal *NeuroImage* over the past decade. We create a network in which nodes represent research topics, and edges give the degree to which these topics tend to be covered in tandem. The network displays small-world architecture, with communities characterized by common imaging modalities and medical applications, and with hubs that integrate these distinct subfields. Using node-level analysis, we quantify the structural roles of individual topics within the neuroimaging landscape, and find high levels of clustering within the structural MRI subfield as well as increasing participation among topics related to psychiatry. The overall prevalence of a topic is unrelated to the prevalence of its neighbors, but the degree to which a topic becomes more or less popular over time is strongly related to changes in the prevalence of its neighbors. Finally, we incorporate data from *PNAS* to investigate whether it serves as a trend-setter for topics' use within *NeuroImage*. We find that popularity trends are correlated across the two journals, and that changes in popularity tend to occur earlier within *PNAS* among growing topics. Broadly, this work presents a cohesive model for understanding the emergent relationships and dynamics of research topics within *NeuroImage*.

1. Introduction

In many fields of research, scientists develop intuitive knowledge of which topics are popular, which might be on the horizon, and which tend to be studied in tandem. Yet each scientist's view of the research landscape is based on a subsampling of the full space, depending on the nature and extent of their experiences in the field. It is therefore often daunting for those who are new to a field to construct even a superficial picture of the research landscape. Moreover, even for those scientists that are steeped in a particular research area, it can be challenging to imagine new connections that might be drawn between topics that historically have been thought of as unrelated.

Recently, the emerging field of network science has proven useful for gaining an understanding of the broader space of scientific research (Fortunato et al., 2018). Previous work on collaboration and citation networks has provided insight into authors social patterns (Newman, 2001a; Borrett et al., 2014; ContandriopoulosArnaud et al., 2016), important studies and turning points in the literature (Chen, 2004), and the large-scale structure of the scientific landscape (Newman, 2001b; Wallace et al., 2012). When examining networks of researchers and networks of articles, the scientific landscape has generally been found to show small-world properties, reflecting clustering within specialty and efficient paths between specialties (Newman, 2001b; Wallace et al., 2012). Yet individual fields show variability in their specific network topology (Newman, 2001b, 2004), opening the door for greater understanding of how any given field efficiently carries out and disseminates scientific research.

In the field of neuroimaging, these types of bibliometric approaches have recently gained popularity. In particular, recent work has studied the most impactful neuroimaging papers (Kim et al., 2016), revealed text-based subfields within functional neuroimaging and their relationships with the activation of specific brain regions (Alhazmi et al., 2018),

* Corresponding author.
E-mail address: jdwor@pennmedicine.upenn.edu (J.D. Dworkin).

https://doi.org/10.1016/j.neuroimage.2018.09.005
Received 8 June 2018; Received in revised form 31 August 2018; Accepted 2 September 2018
Available online 6 September 2018
1053-8119/

and created networks that characterize the relationships between research topics within cognitive neuroscience (Beam et al., 1949). Unlike the study of co-authorship and citation networks, this latter study instead uses a technique that quantifies the relationships between scientific ideas. Here, the operationalization of science as a set of interconnected ideas provides a unique opportunity to study how research topics are related within and across sub-disciplines, and how these topics and their relations grow and change over time.

As neuroimaging researchers, we sought to apply this technique to literature from our field, and to use this framework to simultaneously investigate large-scale and node-scale network features both overall and over time. In this work, we apply graph theoretical approaches to a network of the 100 most common topics covered in the journal *NeuroImage* over the ten-year span from 2008 to 2017. We discuss the large-scale structure of this network, the communities of research areas that emerge from the topic relationships, the roles of individual topics in shaping the network, the ways in which these roles have changed over time, and the potential network and literary foundations of topic popularity. In sum, our study offers unique insights into the nature and use of scientific research in contemporary neuroimaging.

2. Methods

2.1. Data collection

For this study, we retrieved keywords and abstracts from 8547 articles published in *NeuroImage* between 2008 and 2017. We used the keyword sections to create a list of potential topics to be searched for in the abstracts. We chose this technique over latent topic modeling for two reasons: (1) it reflected scientists' explicit opinions as to the words and phrases that constitute relevant scientific topics, and (2) it allowed for the incorporation of multi-word phrases.

To develop a list of potential network nodes, we manually curated the list of topics. The specific curation procedure that we implemented was constructed so as to address potential sources of variation in the topics. First, variability in how researchers referred to topics was manually adjudicated, and different terminology for the same idea was consolidated. For example, *functional magnetic resonance imaging* and *functional magnetic resonance images* were considered to be referring to the same topic. Second, common abbreviations were detected by linking multi-word phrases to their associated parentheticals in keyword and abstract text. Moreover, all variations of the full phrase were replaced by the relevant abbreviation in the abstract and keyword text. For example, variations of the topic *functional magnetic resonance imaging* were found to be associated with the abbreviation *fMRI*, and we therefore replaced references to these terms by *fMRI* in the abstract and keyword text.

2.2. Network construction

We calculated the prevalence of each potential topic by finding the proportion of abstracts or keyword sections that contained at least one reference to the given topic within the timespan of study. We used the 100 most common topics between 2008 and 2017 as nodes to construct the network. Notably, we chose this value because it represented the approximate number at which the least prevalent words occurred sufficiently often to produce a statistically reliable signal in both static and temporal analyses of the network. To ensure that our findings were not unduly dependent on this choice, we also varied the number of topics chosen to construct the network. The effects of network size on the inferred topology are shown in Table S1.

Edges were weighted by the ϕ coefficient for binary association (Ernest, 1991), representing the degree to which two topics tended to be discussed in the same articles. We applied a threshold of positive significance, removing negative edges and non-significant edges. This step was taken to increase the interpretability of the inter-topic links, leading them to signify a meaningful association between two topics within the neuroimaging literature. Nevertheless, to ensure that our findings were not unduly dependent on this choice, we also performed sensitivity analyses in which we maintained all edges. We report the effects of this choice on the community structure of the network in Fig. S2, and we also discuss those results in a later section.

2.3. Network structure

To quantify the structural features of the full network, we sought to investigate the degree to which topics tended to form tightly connected clusters, as well as the overall level of integration of research topics across the network.

Local topic clustering can be quantified using the **clustering coefficient**, which is defined for a node as the probability that two of its adjacent nodes are connected to one another. The version of the clustering coefficient used here is a measure of transitivity defined as follows (Barrat et al., 2004):

$$c_i^w = \frac{1}{s_i(k_i-1)} \sum_{h,j \in N} \frac{(w_{ij}+w_{ih})}{2} a_{ij}a_{ih}a_{hj}, \tag{1}$$

where N is the set of all nodes, and s_i is the node's **strength**, or the sum of all edge weights originating at node i. The variable k_i is the node's **degree**, or the number of edges originating at node i. Finally, w_{ij} is the edge weight connecting node i and node j, and a_{ij} is 1 if $w_{ij} > 0$ and 0 otherwise. The overall clustering behavior of the network can be obtained by taking the average clustering coefficient over all nodes (Rubinov and Sporns, 2010).

Integration across the network can be quantified using the **characteristic path length** of a network. Path length is defined as the average shortest path length between all node pairs (Watts and Strogatz, 1998). A version of the path length for a weighted network can be defined as follows:

$$L = \frac{1}{n(n-1)} \sum_{i \neq j} d_{ij}, \tag{2}$$

where n is the number of nodes and d_{ij} is the shortest weighted path length between nodes i and j, defined as the inverse of the edge weight w_{ij} and obtained using the algorithm given by Johnson (Donald, 1977).

Notably, these two measures of clustering coefficient and path length can be combined to obtain the **small-world propensity** of a network, which represents the degree to which a network shows similar clustering to that of a lattice network, and similar path length to that of a random network (Muldoon et al., 2016). This metric is similar to the commonly used small-world index (Watts and Strogatz, 1998), σ, but has been shown to be unbiased even in the context of networks with varying densities (Muldoon et al., 2016). Both measures broadly represent how well a network can be characterized as having both disparate clusters and strong between-cluster integration. The small-world propensity of a network is defined as follows:

$$\phi = 1 - \sqrt{\frac{\Delta_C^2 + \Delta_L^2}{2}}, \tag{3}$$

where

$$\Delta_C = \frac{C_{lattice} - C_{observed}}{C_{lattice} - C_{random}} \tag{4}$$

and

$$\Delta_L = \frac{L_{observed} - L_{random}}{L_{lattice} - L_{random}}, \tag{5}$$

with C representing the network clustering coefficient, defined as the