Connecting Dream Networks Across Cultures

Onur Varol*

Filippo Menczer

Center for Complex Networks and Systems Research
School of Informatics and Computing, Indiana University, Bloomington, USA

ABSTRACT

Many species dream, yet there remain many open research questions in the study of dreams. The symbolism of dreams and their interpretation is present in cultures throughout history. Analysis of online data sources for dream interpretation using network science leads to understanding symbolism in dreams and their associated meaning. In this study, we introduce dream interpretation networks for English, Chinese and Arabic that represent different cultures from various parts of the world. We analyze communities in these networks, finding that symbols within a community are semantically related. The central nodes in communities give insight about cultures and symbols in dreams. The community structure of different networks highlights cultural similarities and differences. Interconnections between different networks are also identified by translating symbols from different languages into English. Structural correlations across networks point out relationships between cultures. Similarities between network communities are also investigated by analysis of sentiment in symbol interpretations. We find that interpretations within a community tend to have similar sentiment. Furthermore, we cluster communities based on their sentiment, yielding three main categories of positive, negative, and neutral dream symbols.

Categories and Subject Descriptors

H.3.1 [**Content Analysis and Indexing**]: Linguistic processing; H.3.3 [**Information Search and Retrieval**]: Clustering; Information filtering; G.2.2 [**Graph Theory**]: Graph algorithms; Hypergraphs; Network problems

Keywords

Dream interpretation networks; multicultural network; interconnected networks; network communities; oneirology.

*Corresponding author: ovarol@indiana.edu

WWW'14 Companion, April 7–11, 2014, Seoul, Korea.
ACM 978-1-4503-2745-9/14/04.
http://dx.doi.org/10.1145/2567948.2580073.

1. INTRODUCTION

Systems in society and nature contain connections that capture useful information and provide insights for our understanding. In particular, representing knowledge as a network offers depth to our understanding of nature and society. Components of these systems can be mapped as a network to analyze. For example networks of drugs [36] and diseases [13] help us to design new drugs and learn more about human diseases. Networks of products are also studied for modeling markets and the economy [17]. Using networks to model brain connectivity helps to understand functional parts of brains and mechanisms of these functional regions [7]. Networks are also useful for understanding human mobility, social relations, political discussion [8], human dynamics, and society on a large scale [6, 33, 10]. As network science evolves, it also can be seen in novel areas such as cuisine, where it is being used to characterize the eating habits of different cultures [1], and science of science, where it contributes to our understanding of the emergent relationships between disciplines, scholars, and papers [32].

Interconnected networks (with multiple connected layers) recently have attracted the interest of the scientific community. Researchers have investigated statistical mechanics [2] and phase transitions of multiplex networks [27] (where there is a 1-1 correspondence between nodes in different layers), diffusion dynamics [14, 29], and emergence of these networks [25]. Various networks in society can be represented as multiplex networks. Our goal is to advance *oneirology* (the scientific study of dreams) by building an interconnected network of dream interpretations.

In spite of the universal and common experience of dreaming, its purpose and mechanisms are still largely unknown. One of the earliest scientific studies in oneirology dates from about 300 years ago [28]. Throughout history people have recorded dreams in an attempt to explain their connection with waking life [15]. Content analysis of dreams was studied by psychologists Sigmund Freud, Carl Jung, and Calvin Hall [12, 19, 15]. While Freud thought dreams to be aimed at unconscious fulfillment and explained symbols through metaphor, Jung argued that the purpose of dreams were unconscious messages to the self. Calvin Hall collected more than 50,000 dream reports and annotated more than 1,000 of them. He found that dreams all over the world contain similar concepts: the self, other people, and situations [16]. Recently codebooks for dream interpretation and collections of dreams have become available online in the DreamBank database.[1] Dream contents have been analyzed to inves-

[1] http://DreamBank.net

tigate similarities between dreams and waking life, observing common terms about religion and sexuality [9]. Schweickert [30] built and compared networks of individuals in dreams and waking life.

In this study, we collect online data from dream dictionaries and analyze it using network analysis techniques to unveil communities in dream interpretation networks and associations between symbols.

1.1 Contributions and outline

In the remainder of the paper we make the following contributions:

- We discuss how we collected and built dream interpretation datasets for three different languages as introduced in § 2.1 and create the first dream interpretation network as described in § 2.2.
- In § 2.3 we show how to build an interconnected multicultural dream network by using dictionary data for dream symbols.
- § 3.1 reports on community detection in dream interpretation networks, and on our analysis of the characteristics of the resulting communities. The central nodes in these communities are inspected to grasp their topics, finding that representative nodes in the same community convey similar messages.
- In § 3.2 we measure node strength and edge weight correlations between layers, finding that dream symbols in different languages have positively correlated properties.
- Finally, § 3.3 highlights the role of sentiment in similar communities in the English network and the clustering of communities with similar happiness scores.

2. EXPERIMENTAL SETUP

In this section we describe the methodology we followed to collect three datasets of dream interpretations in different cultures and to construct content similarity networks that allow us to study cultural aspects of dream interpretation.

2.1 Datasets

To build our dream interpretation datasets for different cultures, we crawled websites of online dream dictionaries from various sources. The crawled Web pages were parsed to extract dreams symbols with their associated interpretations.

Pre-processing techniques specific to each language (described next) were applied to the raw data extracted from Web pages in that language. Information retrieval techniques were then used to convert text into vector space representations, used for building dream interaction networks. To study dreams across different cultures, we selected the three languages and sources below.

English.

The dream dictionary that we crawled[2] contains 1,391 distinct symbols (English terms). We populated our dataset with these symbols and their interpretations. Interpretations of symbols were pre-processed using lemmatization, Porter stemming [26], and removal of stop-words.

[2] `http://www.dreamdictionary.org`

Chinese.

We crawled the content of a website[3] for traditional Chinese dream interpretations and collected the interpretations of about 1,140 distinct symbols. Unlike many languages, Chinese is written without using spaces to separate words. In Chinese text retrieval, segmentation is an important preprocessing step. We used software for segmentation of Chinese text[4] and removed unicode punctuation characters in Chinese.

Arabic.

We collected dream symbols and interpretations in Arabic from a dictionary website[5] containing 2,419 distinct symbols. Arabic retrieval tasks require some effort to clean text data. We used the ISRI Stemmer [34] integrated in the NLTK [3] software package, which enables the removal of stop-words, language-specific dialectics, initial characters, and prefixes.

2.2 Building networks

To investigate relations between dream symbols and make comparisons between different cultures, we built dream interpretation networks in each language. These networks are weighted and undirected. Each node i corresponds to a symbol in a dream dictionary and each weight e_{ij} to the similarity between interpretation documents for symbols i and j, respectively. Higher weights represent similar interpretations and closer meanings.

For the computation of document similarities we employed a commonly used vector space representation, namely the TF-IDF [18] vector d_i for the interpretation document of each symbol i. Similarities between document representations are computed by the cosine similarity

$$e_{ij} = \frac{\sum_{w \in W_i \cap W_j} d_{iw} d_{jw}}{\sqrt{\sum_{w \in W_i} d_{iw}^2} \sqrt{\sum_{w \in W_j} d_{jw}^2}}, \tag{1}$$

where d_{iw} is the TF-IDF weight assigned to term w in document vector d_i and W_i is the set of words in d_i.

To analyze the resulting weighted networks, we applied a multi-scale backbone extraction technique [31] to remove statistically insignificant edges. In this algorithm, the significance level of edges is controlled by a parameter α. We tuned α for each network to obtain a minimal backbone, i.e, a network containing the minimum number of edges such that all nodes are in a single connected component. The structural properties of the resulting networks are summarized in Table 1. The networks are of course very different, however they display some structural similarities. All of them have high density and clustering coefficients, and short path lengths. In other words, they are small-world networks [35].

Fig. 1 displays distributions of a few network properties. Similar characteristics are observed in the networks corresponding to different languages. The distributions of degree, strength (weighted degree), and edge weights are narrow (Poissonian), while text length has a skewed distribution spanning several orders of magnitude.

[3] `http://zgjm.xixik.com`
[4] `https://github.com/fxsjy/jieba`
[5] `http://dreams.svalu.com`

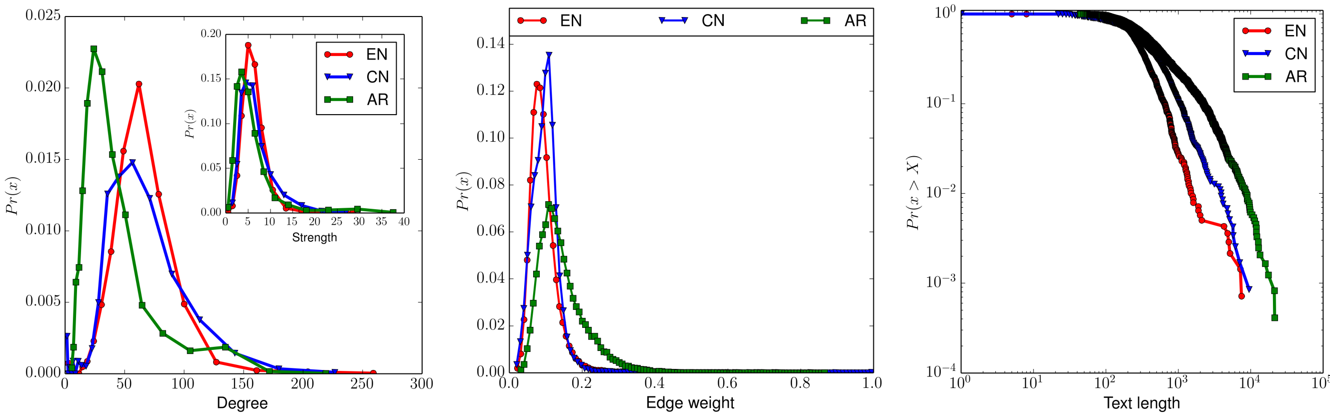

Figure 1: Distributions of network properties: (a) degree and strength (inset), (b) edge weight, and (c) textual length of dream symbol interpretations (CDF).

Table 1: Structural properties of dream interpretation networks generated by backbone extraction for appropriate values of the parameter α.

Network	English ($\alpha = .05$)	Chinese ($\alpha = .15$)	Arabic ($\alpha = .05$)
# nodes	1391	1140	2419
# edges	46302	39864	55893
Density	0.048	0.061	0.012
Avg. degree	66.57	69.94	46.21
Avg. strength	6.26	6.91	6.68
Clustering coefficient	0.15	0.21	0.17
Avg. shortest path	2.05	2.05	2.51

2.3 Multicultural dream network

Building interconnected networks requires a set of nodes appearing in multiple dream interpretation networks. Direct translation is the easiest method for matching nodes in different networks. We used Google's translation service to translate Arabic and Chinese symbols into English, and employed the English terms to match symbols in different layers.

3. RESULTS

In this section we carry out community detection for the dream interpretation networks and compute correlations between multiple layers. We observe correlations that points to similarities between different cultures. Finally, we explore the role of emotions in dream interpretations by performing sentiment analysis in the English network.

3.1 Communities

Communities provide a more compact representation of large networks [11]. The identification of the central nodes in a community allows us to recognize the role or meaning of the community.

We detected communities in dream interpretation networks by using the Louvain algorithm [4], as implemented by Guillaume [22], which performs a greedy optimization of Newman's modularity [24] on a weighted network. The Louvain method was executed 500 times for each network to find the best modularities: 0.24, 0.37, and 0.44 for English, Chinese, and Arabic, respectively.

Since the networks are built from symbol similarities, it is natural to expect that a community should join related symbols corresponding to a shared topic. To identify topics, we extract the central nodes in each community using the eigenvector centrality algorithm, which gives credit to a node by considering both its degree and that of its neighbors [5]. Fig. 2 depicts coarse-grained networks in which nodes represent the communities. An edge between two community nodes has a weight equal to the total weight of the links connecting symbols in the two communities.

According to Hall's cognitive theory of dream symbols [15], we expected to observe communities containing dream symbols that serve either as warning for the self or reactions to daily life and others. Automatic translation of the symbols helps to determine if these meanings are preserved across networks.

In the English network (Fig. 2(a)), the community that contains 'goal,' 'hill,' and 'ladder' is interpreted as achievement after a struggle, while the community labeled 'voices,' 'mailman,' and 'message' represents warning and precaution. The community with symbols 'hiding,' 'guard,' and 'raincoat' means having protection.

The Chinese community in Fig. 2(b) that contains the symbols 'rainbow,' 'poor,' and 'quarrel' tells about good things that will happen in the future, while the one with the nodes 'road accident,' 'woman,' and 'illusion' contains symbols about forewarning and need of life changes. The community labeled 'suicide,' 'walking,' and 'enemies' is interpreted as describing happy life and end of troubles. Note the rough correspondence between the first two example communities mentioned for English and Chinese networks.

In the Arabic network represented in Fig. 2(c), the role of religious characters and objects are identifiable in distinct communities represented in purple and red. Similarly in the Chinese network, the yellow community contains symbols like 'god' and 'vision.' Content analysis of symbols in DreamBank leads us to similar conclusions about religious dreams [9].

Given the low reliability of automatic translation, we do not want to make any strong claims about correspondence between interpretations in different cultures. However, we

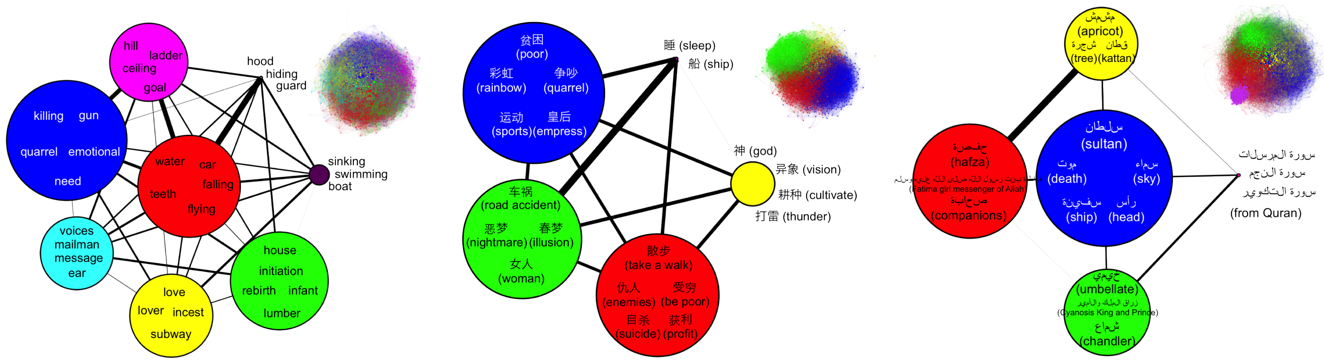

Figure 2: Communities detected by Louvain method: (a) English, (b) Chinese, and (c) Arabic. To label each community, we list the symbols with the highest Eigenvector centrality. The complete networks are illustrated in the upper right corners of each figure, with colors corresponding to the communities (colors are not intended to match across different networks).

can compare the networks by investigating correlation between strengths and edge weights.

3.2 Correlation between layers

One way to look at the structural similarities between dream interpretation networks in different languages is to focus on common nodes (detected by automatic symbol translation) and measure the correlations between properties of these nodes. Let's consider two properties: strength of common nodes and weight of edges connecting pairs of common nodes. We measure Pearson's correlation r. For strength the definition is straightforward. For edge weight the correlation is defined as follows:

$$r(A,B) = \frac{\sum\limits_{i,j \in S^A \cap S^B} (e_{ij}^A - \overline{e^A})(e_{ij}^B - \overline{e^B})}{\sqrt{\sum\limits_{i,j \in S^A} (e_{ij}^A - \overline{e^A})^2} \sqrt{\sum\limits_{i,j \in S^B} (e_{ij}^B - \overline{e^B})^2}},$$

where $\overline{e^A}$ is the average weight of edges in network A. We only consider edges between nodes common to both networks. S^A is the set of edges among common nodes that are present in A.

Table 2 reports on the numbers of common nodes and edges, as well as correlations between layers. The correlations are not very strong, but positive and significant in almost all cases. These results indicate that interconnected nodes have somewhat similar connectivity: symbols tend to be connected to the same neighbors across languages. Symbols also tend to have similar strength: general (specific) symbols in one language tend to be general (specific) in other languages as well. These results hold whether we consider the backbone or fully connected networks.

3.3 Sentiment analysis

Sentiment analysis is an important tool for probing textual data to identify nuanced differences, such as happiness. In dreams, emotions and feelings play important roles in the interpretation of the meaning of symbols. We analyzed the symbol interpretations in each community to study whether communities can be further grouped into larger clusters based on similar sentiment. We computed a happiness score for each symbol using a dataset of happiness scores in English [20]. This dataset contains happiness scores for more than 10,000 English words, each ranging between 1 and 9 (1: least happy, 5: neutral, and 9: most happy). Each interpretation text contains a set of words that has a happiness score in the dataset. The happiness score for the symbol is averaged across the words in the interpretation text.

We further investigated the sentiment properties of the dream network. Building a sentiment similarity network among symbols is infeasible, because we cannot evaluate statistical significance of the similarity between two individual sentiment scores. Therefore, we focused on the relationships between sentimentally similar communities. The symbols in each community produce happiness score samples drawn from the distribution of sentiment. These samples allow us to compute statistical significance (p-values) of differences between community distributions, using two-sample t-tests. The p-values are then used as similarity measures between communities, and passed to an aggregative hierarchical clustering algorithm to join sentimentally similar communities. Fig. 3 reports the similarities between communities and illustrates the dream clusters with similar sentiment. A multiple comparison Kruskal-Wallis test [21] of sentiment scores confirms that the clusters are highly significant. The results are not particularly dependent on the similarity measure; Jensen-Shannon divergence [23] between the distributions yields similar partitions.

The most central symbols in the sentimentally similar communities are illustrated in Table 3. Each community in the English dream interpretation network deserves investigation. Interpretations of symbols in C1 concern fresh starts and new beginnings; C4 mostly deals with warnings about upcoming life changes and suggests awareness to surrounding events. C7 contains symbols that describe forms of protection. These three communities concern change, forewarning, and protection from the unexpected. Similarly, we analyzed the symbols in communities C2 and C5, which have the lowest sentiment scores. These symbols deal with troubles in life and personal relationships and encountering highly stressful situations. The last group of communities has the highest sentiment scores. Symbols in communities C0 and C3 are associated with strong emotions about specific situations and others. C6 contains symbols dealing with ambition for reaching targets or achieving success.

Table 2: Pearson correlation for pairs of common nodes (interconnections) identified by translation of symbols. We compute the correlation of edge weight and node strength over two types of network: fully connected and backbone. The correlation values that are statistically significant ($p < 0.05$) are marked in bold, based on the sample sizes (common nodes or edges).

Networks	Common nodes	Common edges		Node strength r		Edge weight r	
		Backbone	Full	Backbone	Full	Backbone	Full
Arabic & Chinese	378	107	68,162	**0.09**	0.06	**0.24**	**0.04**
Chinese & English	274	151	32,196	**0.20**	**0.13**	**0.15**	**0.05**
Arabic & English	239	39	26,193	0.10	**0.23**	**0.43**	**0.12**

In all of these clusters of communities, sentiment plays a meaningful and binding role. Inspection of the central nodes in the Arabic and Chinese network communities suggests roughly similar relations, in agreement with Hall's cognitive theory on dream symbols [15].

4. CONCLUSIONS

In this work we introduced the concept of a dream interpretation network. Our study bridges network science and oneirology; properties of communities were investigated using network science and sentiment analysis techniques. Our analysis was carried out on multiple networks in different languages to investigate the role of culture in the associations of dream symbols. We sought to understand communities in dream interpretation networks, and our findings support Hall's cognitive theory of dreams.

We built a multicultural dream network by identifying interconnections between language layers. Nodes with interconnections are sparse compared to the overall network due to our reliance on automatic translation to match symbols in different languages. We also explored the identification of interconnections by using a Thesaurus Web service available online.[6] Such an approach did not improve the correlations between layers, possibly because this Web service does not sort synonyms by their confidence. In the future, it would be interesting to investigate cross-cultural relations by using other online resources, such as Wikipedia.

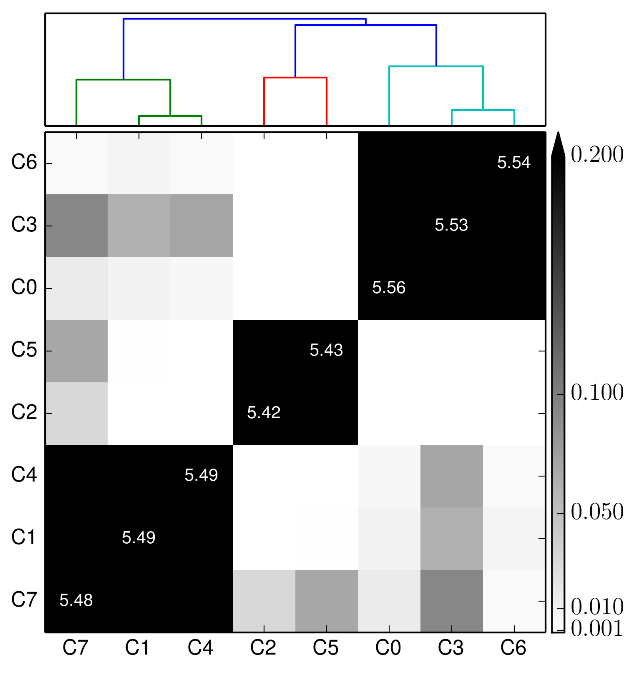

Figure 3: Hierarchical clustering of communities in the English dream interpretation network, using symbol happiness distributions to compute similarities (darker means more similar). The values on the diagonal represent average happiness scores within each community.

Table 3: Communities in the English dream interpretation network, grouped by their sentiment similarities. The groups are identified by hierarchical clustering. The symbols with the highest eigenvector centrality are used to label each community.

Community	Central symbols
C1	house, rebirth, initiation, infant, lumber
C4	voices, mailman, message, fax machine, ear
C7	hiding, overcoat, guard, hood, raincoat
C2	falling, car, teeth, flying, water
C5	gun, need, killing, quarrel, emotional
C0	houseboat, boat, sinking, swimming, scuba diving
C3	love, lover, subway, love note, incest
C6	goal, ceiling, hill, ladder, rock climbing

A more detailed analysis with different algorithms of community detection in the multilayer symbols network is left for future studies. Symbols such as 'water,' 'teeth,' and 'flying' have multiple explanations depending on actors or associated colors and other context. They will require extensive analysis for deeper understanding. Content analysis of user-reported dreams have been carried out to identify words associated with diverse emotions [9]. It would also be desirable to study the co-occurence of symbols in dreams to derive a better understanding of associations between symbols.

5. ACKNOWLEDGMENTS

We are grateful Yong-Yeol Ahn, the NaN group (`cnets.indiana.edu/groups/nan`), and three anonymous reviewers for helpful feedback, and to John McCurley for editing assistance. This work is partially supported by DARPA (grant W911NF-12-1-0037), NSF (grant CCF-1101743), and the McDonnell Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

[6] `http://thesaurus.altervista.org`

6. REFERENCES

[1] Yong-Yeol Ahn, Sebastian E Ahnert, James P Bagrow, and Albert-László Barabási. Flavor network and the principles of food pairing. *Scientific Reports*, 1, 2011.

[2] Ginestra Bianconi. Statistical mechanics of multiplex networks: Entropy and overlap. *Physical Review E*, 87(6):062806, 2013.

[3] Steven Bird, Ewan Klein, and Edward Loper. *Natural language processing with Python*. O'Reilly, 2009.

[4] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, 2008.

[5] Phillip Bonacich. Some unique properties of eigenvector centrality. *Social Networks*, 29(4):555–564, 2007.

[6] Dirk Brockmann, Lars Hufnagel, and Theo Geisel. The scaling laws of human travel. *Nature*, 439(7075):462–465, 2006.

[7] Ed Bullmore and Olaf Sporns. The economy of brain network organization. *Nature Reviews Neuroscience*, 13(5):336–349, 2012.

[8] Michael Conover, Jacob Ratkiewicz, Matthew Francisco, Bruno Gonçalves, Alessandro Flammini, and Filippo Menczer. Political polarization on twitter. In *Proc. 5th International AAAI Conference on Weblogs and Social Media (ICWSM)*, 2011.

[9] G William Domhoff and Adam Schneider. Studying dream content using the archive and search engine on dreambank. net. *Consciousness and Cognition*, 17(4):1238–1247, 2008.

[10] Emilio Ferrara, Onur Varol, Filippo Menczer, and Alessandro Flammini. Traveling trends: social butterflies or frequent fliers? In *Proc. of 1st ACM Conference on Online Social Networks (COSN)*, pages 213–222. ACM, 2013.

[11] Santo Fortunato. Community detection in graphs. *Physics Reports*, 486(3):75–174, 2010.

[12] Sigmund Freud. *The interpretation of dreams*. Kessinger Publishing, 2004.

[13] Kwang-Il Goh, Michael E Cusick, David Valle, Barton Childs, Marc Vidal, and Albert-Laszlo Barabasi. The human disease network. *Proceedings of the National Academy of Sciences*, 104(21):8685–8690, 2007.

[14] S. Gomez, A. Diaz-Guilera, J. Gomez-Gardenes, C. J. Perez-Vicente, Y. Moreno, and A. Arenas. Diffusion dynamics on multiplex networks. *Phys. Rev. Lett.*, 110:028701, Jan 2013.

[15] Calvin S Hall. A cognitive theory of dream symbols. *The Journal of General Psychology*, 48(2):169–186, 1953.

[16] Calvin S Hall and Robert L Van de Castle. *The content analysis of dreams*. Appleton-Century-Crofts, 1966.

[17] César A Hidalgo, Bailey Klinger, A-L Barabási, and Ricardo Hausmann. The product space conditions the development of nations. *Science*, 317(5837):482–487, 2007.

[18] Karen Sparck Jones. Index term weighting. *Information Storage and Retrieval*, 9(11):619–633, 1973.

[19] Carl G Jung. Memories, dreams. *Reflections*, 84, 1963.

[20] Isabel M Kloumann, Christopher M Danforth, Kameron Decker Harris, Catherine A Bliss, and Peter Sheridan Dodds. Positivity of the english language. *PloS One*, 7(1):e29484, 2012.

[21] William H Kruskal and W Allen Wallis. Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47(260):583–621, 1952.

[22] Etienne Lefebvre and Jean-Loup Guillaume. Louvain method: Finding communities in large networks. `sites.google.com/sites/findcommunities`, 2008.

[23] Jianhua Lin. Divergence measures based on the shannon entropy. *IEEE Transactions on Information Theory*, 37(1):145–151, 1991.

[24] Mark EJ Newman and Michelle Girvan. Finding and evaluating community structure in networks. *Physical Review E*, 69(2):026113, 2004.

[25] V. Nicosia, G. Bianconi, V. Latora, and M. Barthelemy. Growing multiplex networks. *Phys. Rev. Lett.*, 111:058701, Jul 2013.

[26] Martin F Porter. An algorithm for suffix stripping. *Program: Electronic library and information systems*, 14(3):130–137, 1980.

[27] Filippo Radicchi and Alex Arenas. Abrupt transition in the structural formation of interconnected networks. *Nature Physics*, 2013.

[28] Hervey De Saint-Denys. *Dreams and How to Guide Them*. Duckworth Pub, 1982.

[29] Anna Saumell-Mendiola, M Angeles Serrano, and Marian Boguna. Epidemic spreading on interconnected networks. *Physical Review E*, 86(2):026106, 2012.

[30] Richard Schweickert. Properties of the organization of memory for people: Evidence from dream reports. *Psychonomic Bulletin & Review*, 14(2):270–276, 2007.

[31] M Ángeles Serrano, Marián Boguñá, and Alessandro Vespignani. Extracting the multiscale backbone of complex weighted networks. *Proceedings of the National Academy of Sciences*, 106(16):6483–6488, 2009.

[32] Xiaoling Sun, Jasleen Kaur, Stasa Milojevic, Alessandro Flammini, and Filippo Menczer. Social dynamics of science. *Scientific Reports*, 3(1069), 2013.

[33] Michael Szell, Renaud Lambiotte, and Stefan Thurner. Multirelational organization of large-scale social networks in an online world. *Proceedings of the National Academy of Sciences*, 107(31):13636–13641, 2010.

[34] Kazem Taghva, Rania Elkhoury, and Jeffrey Coombs. Arabic stemming without a root dictionary. In *International Conference on Information Technology: Coding and Computing*, volume 1, pages 152–157. IEEE, 2005.

[35] Duncan J Watts and Steven H Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, 1998.

[36] Muhammed A Yıldırım, Kwang-Il Goh, Michael E Cusick, Albert-László Barabási, and Marc Vidal. Drug target network. *Nature Biotechnology*, 25(10):1119, 2007.