Journal of Informetrics 12 (2018) 893–905

Contents lists available at ScienceDirect

Journal of Informetrics

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

Topology-driven trend analysis for drug discovery

Yanhua Lv[a], Ying Ding[b,c], Min Song[d,*], Zhiguang Duan[a,e]

[a] School of Management, Shanxi Medical University, 56 Xinjian South Road, Taiyuan, Shanxi, China
[b] School of Informatics & Computing, Indiana University, 107 S. Indiana Ave, Bloomington, IN, USA
[c] School of Information Management, Wuhan University, Wuhan, Hubei, China
[d] Department of Library and Information Science, Yonsei University, Seoul, South Korea
[e] Shanxi University of Chinese Medicine, 121 Daxue Street, Jinzhong, Shanxi, China

ARTICLE INFO

Article history:
Received 12 March 2018
Received in revised form 23 July 2018
Accepted 23 July 2018

Keywords:
Network analysis
Literature-based discovery
Text mining

ABSTRACT

The primary goal of the present study is to discover new drug treatments by topology analysis of drug associations and their therapeutic group network. To this end, we collected 19,869 papers dated from 1946 to 2015 that are related to autism treatment from PubMed. We extracted 145 drugs based on MeSH terms and their synonyms (the total number is 6624) within the same ATC classification hierarchy and used them to find drug associations in the collected datasets. We introduced a new topology-driven method that incorporates various network analyses including co-word network, clique percolation, weak component, pathfinding-based analysis of therapeutic groups, and detection of important drug interaction within a clique. The present study showed that the in-depth analysis of the drug relationships extracted from the literature-based network sheds new light on drug discovery research. The results also suggested that certain drugs could be repurposed for autism treatment in the future. In particular, the results indicated that the discovered four drugs such as Tocilizumab, Tacrolimus, Prednisone, and Sulfisoxazole are worthy of further study in laboratory experiments with formal assessment of possible effects on symptoms, which may provide psychologists, physicians, and researchers with data-based scientific hypotheses in autism-drug discovery.

1. Introduction

Trajectory analysis of the research profile of successful therapeutic drugs is of paramount importance for the drug research community due to the following reasons. First, by analyzing research trends and thrusts of drugs, the evolution of the research profile of drugs can be understood, which in turn enables to scrutinize the impact of the initial use of the drugs in experimental studies on the subsequent studies of the drugs used for drug to humans. Second, inference of the new relationship between drugs throughtime-series analysis leads to a new drug discovery. However, despite the cruciality of such needs and potential synergetic contributions of informetics approaches, informetics has not been fully applied to solve the core problems of drug discovery. The majority of informetrics research in drug study dealt with measuring scholarly activities of drug discovery either by co-word or co-MeSH term analysis (Bordons, Bravo, & Barrigón, 2004; Hong et al., 2016; Leydesdorff, D. Rotolo, & I., 2012). Otherwise, bibliometrics approaches to drug discovery were mainly focused on how particular public databases or resources have an impact on drug discovery (Cheng, Pan, & Hao, 2014). To the best of our knowledge, there has been no

* Corresponding author.
E-mail address: min.song@yonsei.ac.kr (M. Song).

https://doi.org/10.1016/j.joi.2018.07.007
1751-1577/

previous study of applying the informetrics approach for solving new drug discovery. The proposed approach is the first attempt to identify new relationships among therapeutic groups of drugs and use them to infer new drug discovery for autism treatment by mapping drug associations onto their therapeutic groups. Our informetrics-centric approach coupled with text mining makes it possible to make use of trajectory analysis of therapeutic group networks for new drug discovery.

As the subject for the present study, we chose autism. Despite the increasing diagnosis of autism over the last twenty years and extensive biomedical research on brain and nervous system disorders, and related pharmacological problems, there has been little progress in the development of pharmacological treatments for the social impairments that are at the core of this disorder (FDA, 2009; Martin, Scahill, Klin, & Volkmar, 1999; Modi & Young, 2012). Accordingly, Norén argued that data-driven discovery can be considered in drug repositioning because it can produce no pre-specified hypothesis (Norén, 2011). As asserted by Norén, the large amount of pharmacological and biological knowledge available in the literature makes it an increasingly feasible to find novel drug indications for existing drugs using an in silico approach. By combining network analysis with text mining, the goal of the present study is two-fold: 1) to identify research trends pertinent to autism treatments, by discovering novel interactions among different therapeutic groups in autism research; 2) to identify drug repurposing opportunities for autism and proposing possible new scientific hypotheses by novel interactions among drugs discovered from step 1.

A series of experiments showed that together with text mining techniques, our proposal of the topology-driven analysis was able to detect the new therapeutic groups of drugs and new, plausible drug discovery in autism treatment research. In particular, the present study revealed that 50% of research focused on drugs in the same therapeutic group in the early stage of autism research (prior to mid-2000s). However, this proportion decreased with time, and more than 70% of research focused on cross-therapeutic group drugs in the past 10 years. In addition, the results of the study identified that core therapeutic groups of autism have steadily changed over time. However, there is a stable subgroup existing in the drug network (i.e., that never changes with time in structure), which forms the foundation in autism treatment. In the drug network, we identified four drugs as worthy of further study in laboratory experiments, which may provide psychologists, physicians, and researchers with data-driven scientific hypotheses in autism-drug discovery.

2. Related work

All drugs used in clinical medicine require large-scale trials before approval (Jeong, Heo, Kang, Yoon, & Song, 2016), which provides a wealth of material in the published literature about the biological activities and safety of the drugs. However, the amount of such information is now too large for any one person to keep abreast of, even in a niche area of research. Scientific publications are growing at an exponential rate (Larsen & Von Ins, 2010), with over 50 million papers published so far (Jinha, 2010), and over a million additional articles published annually (Björk, Roos, & Lauri, 2008). That means on average a new article is published every 30 s. At the same time, digital publications are narrowing their science and scholarship focus, as well as the range of findings and ideas built upon them (Evans, 2008). Increasing interest in improving treatments for autism also has led to a surge in publications in this field; far more than any individual scientist can keep up with. As a result, data mining and information technologies have proven necessary for good research-based decision-making (Bianchi et al., 2014; Chen, Ding, & Wild, 2012). Articles have been essential to bibliometric studies for decades (Ying et al., 2013). Based on articles, the concept of Entitymetrics was proposed (Health, 2018) to measure the impact of knowledge units at various levels. One kind of entitymetrics are micro-level knowledge entities, such as genes, drugs, and diseases etc., and they act as carriers of knowledge units in scientific articles to identify the importance of entities embedded in scientific literature for further knowledge discovery. In the present research, literature-based knowledge discovery aims to connect the potential relationships of scientific entities to generate new knowledge from the perspective of drugs.

Secondly, as medicine advances, all approved drugs must be novel and important to be reasonably characterized as cures in their own domains. However, drug research and development requires huge investments, which means that an unclear target might result in additional costs or inefficiency (Williams, Lotia, Holloway, & Pico, 2015). From this perspective, a scientific hypothesis can facilitate medical trials, as well as providing substantial data for the advance of science and improvement of medicine. Although experiment-based knowledge discovery is based on stringently validated data from experiments or clinical trials, the benefits of literature-based discovery can be enormous in helping domain experts to form scientific hypotheses. The connections between concepts in scientific literature can be established if two concepts co-occur in a predefined context (e.g., title, abstract, one sentence, or one paragraph), and researchers had verified its valuation. Stegmann and Grohmann also argued that co-occurrence network analysis is a powerful method for literature-based hypothesis generation and knowledge discovery by finding characteristic values in the co-keyword analysis which allow a rapid identification of possible cluster based on centrality-density ratio (Stegmann & Grohmann, 2003). After that, a study shows that Metformin changes the peroxisome proliferator-activated receptor in the uterine tissue of mice (Blumberg et al., 2013), and then researchers identified the interaction between Metformin and the peroxisome proliferator-activated receptor through a bio-entity citation network. In 2013, a research identified of an interaction between Metformin and Resistin, which supported (Newschaffer et al., 2007) the hypothesis that Metformin treatment had a positive impact on up-regulating Resistin.

Given the rapid growth of scientific literature, literature-based approaches to generating hypotheses automatically have gained increasing attention in recent years. Swanson discovered that certain unintended logical connections across scientific domains and potentially revealing of new knowledge, were enabled by reference citations or other bibliographic clues