Temporal Formation and Evolution of Online Communities

Hossein Fani
Faculty of Computer Science, University of New Brunswick, NB, Canada
Laboratory for Systems, Software and Semantics (LS^3), Ryerson University, ON, Canada
hosseinfani@gmail.com

ABSTRACT

Prediction of users' behaviour, interests, and influence are of interest within the realm of online social networks due to the wide range of applications, such as personalized recommendations and marketing campaigns. However, the proposed approaches are not always scalable to a large number of users and huge amount of content generated by them. Community-level studies are introduced to help facilitate scalability, among others, highlighting the main properties of the network at a higher collective macro level.

The state-of-the-art community detection methods mainly focused on identification of online communities and possible impacts or influence relationships from several aspects, e.g. communities of users that are formed based on shared relationships and topological similarities, or communities that consist of users that share similar content. However, there are not much existing research work studying detection of communities that simultaneously share topical and temporal similarities. For example, users who contribute to the topic '*Julian Assange Allegations by Swedish Prosecutors*' might show their inclination with a week time delay due to their interest priority. The community of users who are interested in the topic *this week* could be different from those who become interested in the *following week*. In our proposed method we identify user communities that have similar temporal dispositions according to their topics of interest. To this end, we apply multivariate time series analysis to measure the inter-user similarity. The proposed approach will be able to identify the following types of communities: *i)* the community of users who are only interested in a given topic and in a same time period; therefore, supporting topic-driven community detection. *ii)* the communities of users who are interested in a same topic but in different time periods; hence, supporting temporality in the community detection process.

Furthermore, the existing studies have not explored the *causation* relationship between the communities, named inter-community causation, i.e. a community, the *cause*, makes other community, the *effect*, happen. Causation provides systematic explanation as to why communities are formed and helps to predict future communities. For example, a community around the topic '*Julian Assange and the WikiLeaks*' gives rise to intense discussions around topics such as 'Internet Censorship' and '*Blocking Wikileaks*' in the online social networks and cause a new corresponding communities. Since causality is a temporally asymmetric relation, i.e., the cause *precedes* the effect, we employ Granger concept of causality, G-causality for short, to incorporates precedence through time series. Assuming no hypothesis about the data, Granger's method sits in complete harmony with our community detection model and forms a coherent framework besides its simplicity, robustness, extendability. Our final G-causality-based model will be a weighted directed acyclic graph (DAG) of communities. In this graph, the edge direction represent the cause to effect relationship of its adjacent communities. Figure 1 shows two samples of causation patterns amongst communities.

To sum up, we address two main questions in this research proposal: *i)* how can communities with similar topical and temporal behaviours be identified, and *ii)* how can causation relation between different online communities be detected and modeled. We model users' behaviour towards topics of interest through multivariate time series to identify like-minded communities. Further, we use Granger's concept of causality to infer causation between detected communities from corresponding users' time series. We evaluate the proposed approach empirically on a dataset of tweets. We assess the performance of the proposed community detection methods with the state of the art and verify the causal model through its prediction accuracy.

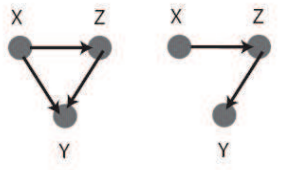

Figure 1: Two possible causation triads

WSDM 2016 February 22-25, 2016, San Francisco, CA, USA

ACM ISBN 978-1-4503-3716-8/16/02.
DOI: http://dx.doi.org/10.1145/2835776.2855089