Network Patterns in South African Election Tweets

Aurona Gerber[1,2(✉)] and Stefanie Strachan[1]

[1] Department of Informatics, University of Pretoria, Pretoria, South Africa
aurona.gerber@up.ac.za
[2] Centre for AI Research (CAIR), Pretoria, South Africa

Abstract. Social communities play a significant role in understanding complex societies, from communities formed by support interactions between friends and family to community structures that depict the flow of information, money and power. With the emergence of the internet, the nature of social networks changed because communities could form disassociated from physical location, and social network analysis (SNA) on social media such as Twitter and Facebook emerged as a distinct research field. Studies suggest that Twitter feeds have a significant influence on the views and opinions of society, and subsequently the formation of communities. This paper reports on a study where social network analysis was performed on Twitter feeds in South Africa around the 2019 elections to detect distinct patterns within the overall network. In the datasets that were analysed, a specific network pattern namely Broadcast Networks were observed. A Broadcast Network typically reflects central hubs such as media houses, political parties or influencers whose messages are repeated without interaction or discussion. Our results indicate that there were few discussions and interactions and that messages were broadcasted from central nodes even though the general experience of Twitter users during this time was of intense discussions and differences in opinion.

Keywords: Social network analysis · Graph analysis · Twitter networks · Community clusters · Network visualisation · South african elections

1 Introduction

How communities form plays a significant role in understanding society and society interactions [1, 2]. In the past decade various studies were done on the usage and influence of the internet, technology, and social media in society [3–5]. These studies found that social media became one of the most important means of communication and therefore forms a significant part of what determines the views and opinions of people [6–9]. As a result, social network analysis (SNA) emerged as a research field. SNA analyse network structures in social media networks, and one of these network structures is networked communities and clusters. Understanding the ways in which online communities form and communicate allows us to interpret the flow of information and opinions, power, money and key influencers [10]. This is particularly valuable when analysing political structures within a specific society, an important

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2021
T. Antipova (Ed.): ICCS 2020, LNNS 186, pp. 3–13, 2021.
https://doi.org/10.1007/978-3-030-66093-2_1

capability because of the recent developments such as large-scale individualized collective action that can be observed globally and that results in substantial turmoil due to protest action [11–13]. Understanding the formation and dynamics of politics in social networks can provide useful insights on the interactions and changes in political communities, and might assist in the design of interventions [11, 12].

The aim of this study was to detect the social network structures that were formed around specific political events within the South African Twitter community, i.e. whether and how communities or groups were formed and organised within the network, and therefore an analysis of the patterns of information flow within and between clusters. The results of this social network analysis could assist with understanding the South African political landscape.

2 Background and Related Work

This section provides an overview of social networks, social network analysis (SNA), and the application of SNA on social media data.

2.1 Social Networks

A social network is a type of complex network and can be described as a social structure composed of a set of social actors or users, and the inter-relations and social interactions between them. These networks are useful to study the relationships between individuals, groups, social units, or societies [14]. The analysis of social networks is centred around the fundamental theory that the social network is made up of the relations and interaction between users and within units rather than by the properties of the user itself [1, 15]. Network structures are detectable in many different contexts, including biological systems, population and cultural structures, and communication networks. One important aspect of analysing social networks is the detection of clusters or crowds made up of users that are closely interconnected. These structural clusters can be analysed by studying both the links within a cluster as well as the links between different clusters [16]. Depending on the research objectives, there are different general analysis levels of social networks, namely micro-level, meso-level, and macro-level [16]. Micro level analysis generally focuses on an individual or small group and expands to trace the social relationships associated with that user or social unit. Meso-level analysis is used for medium sized groups or to determine how connections are formed between micro and macro level systems. Most human social networks fall within the meso-level category. Macro-level analysis is used when the focus is on the outcomes of interactions and communications of a larger network rather than interpersonal relations within the network. Various network analysis approaches can be followed to study the network structures formed by social entities, as well as the possible patterns found within these structures [8]. Social network analysis is used in the identification of network patterns, the exploration of network dynamics, and the study of information flow within and between, specifically, social media networks [8].

2.2 Social Network Analysis

Social network analysis (SNA) is the general reference to the process of investigating social networks or structures within social media through the use of networks, knowledge graphs, and graph theory. A crucial part of analysing any large dataset is the means to represent the information in a meaningful way – specifically when utilizing the results in decision making processes. Visualisation can prove to be an effective way of conveying results in a manner that is easily understood by users [17]. Some of the social structures that commonly use visualisation through social network analysis include social media networks, information sharing, business networks, general social networks, transmission of disease, and association studies. These types of networks are often visualised by focusing on the social links of persons or users, following the same approach as semantic networks where the significance of an entity is determined by its relation to, and interaction with other entities, rather than its own properties. Entities are represented as nodes or points, and the relationships and interactions between entities are represented by interconnecting lines or edges. Depending on the attribute of interest, the nodes and edges can be partitioned into higher-density subgraphs, resulting in the formation of communities [18].

The Social Media Research Foundation [19] has done various studies on social network analysis and the information about individual behaviour, social relationships, and community value that can be extracted. They subsequently released an open-source tool, NodeXL, that implemented most of the network structure algorithms to support social network and content analysis [20]. In a study on the mapping of Twitter topic networks using NodeXL, Smith et al. identified six distinct archetypical network patterns [21, 22]:

- ***Polarized Crowd*** network pattern usually depict a small number of groups depicting highly divisive and polarized discussions with very little connection between the groups. This is typical of groups that don't argue but ignore each other. The distinct groups therefore rely on different information sources and do not interact [21].
- ***Tight Crowd*** network pattern are characterized by "highly interconnected nodes or people with few isolated participants". Such a pattern is typical of communities formed by conferences, or typically groups associated with professional topics or specific hobbies. Such groups therefore support each other with information flows between members of the group [21].
- ***Brand Clusters*** is a network pattern is highly fragmented and is observable when topics such as well-known products, services or celebrities are discussed in Twitter. The nodes tweet about a topic but not to each other, and the larger the population, the less likely it is that participants are connected to one another. Information flows are this about a topic, and not between members of a group, there is therefore not an exchange of ideas, information is just passed on [21].
- ***Community Clusters*** is a network pattern where multiple smaller groups or hubs are formed each with its own audience, influencers, and sources of information. This pattern therefore reflects multiple centres of activity or community each with its own following and information flows, as well as a fair number of isolates. Such a network would typically represent diverse angles on a subject based on its relevance

to different audiences or it could depict a diversity of opinion and perspective on a topic [21].

- ***A Broadcast Network*** pattern is a distinctive hub-and-spoke that typically depict "Twitter commentary around breaking news stories and the output of well-known media outlets" with many people repeating information such as prominent news and media organizations tweets. Members are often connected only to the hub news source, without connecting to one another. Such a network structure would also depict influencers and agenda setters [21].
- ***A Support Network*** pattern is also a hub-and-spoke pattern that is typically observable where customer services for a major business are handled by Twitter service accounts. In the pattern the "hub account replies to many otherwise disconnected users, creating outward spokes" whereas, in the Broadcast pattern, "the hub gets replied to or retweeted by many disconnected people, creating inward spokes". This is a useful pattern for government, businesses, and groups that provide services and support via social media, because it could serve as a benchmark for evaluating performance [21].

Such social network patterns of Twitter communities can be analysed and visualised to provide insight into society, such as power and political sentiment, as well as information flows about the persons and topics that drive conversations and group behaviour.

The social media data source relevant to this study is Twitter. Twitter, an online news and social networking service created in 2006, is a platform dedicated to personal expression through the minimalist concept of microblogging [23]. Posts are limited to 280 characters, aside from a few language exceptions, and users can follow other users with no mandatory interaction. The simplicity makes it an ideal means to report current events and connect around specific topics, people or interests. The amount and availability of the data makes it possible for researchers to analyse the content based on network maps created by tweets, mentions, retweets, and followers.

3 Approach

The aim of this study was to determine whether any distinct patterns could be found within the social network formed through Twitter conversations on South African political issues, and the data was collected around the 2019 elections. Three specific periods were chosen for the dataset around significant political dates, namely, the last week that political campaigning was allowed (28 April 2019–4 May 2019), election day on 8 May 2019, and inauguration day on 24 May 2019. NodeXL was adopted as the toolset for the network analysis. Data was collected through NodeXL supplemented with data sources directly from the Twitter API. The search terms used for collection of the respective datasets were based on the most active hashtags on Twitter during these periods namely:

- Dataset 1 and 2: SAElections2019, SADecides2019, Xse, XseDay, AfterVotingIExpect, AfterIVoteIExpect.
- Dataset 3: PeoplesInauguration, SAInauguration19

A much larger set of unique Tweets was found for dataset 1 and 2 than for dataset 3. Data wrangling included the removal of duplicates determined by the tweeter's username, the mentioned or replied to username if applicable, and the date and time of the Tweet, or the imported id.

All the datasets were imported into NodeXL, and the unique entries translated into graph vertices, with interactions the edges or links. We adopted the Clauset-Newman-Moore algorithm [24] for the detection of clusters or communities. The network was visualised using the NodeXL graph visualisation features. Because it is Twitter datasets, all the graphs were undirected. The top words and hashtags by frequency of mention were determined for the overall network as well as for each group. The connections within and between the different groups can also be analysed to determine how information flows within the network.

Dataset 1 consisted of 9531 unique entries from 4935 unique entities or users. Dataset 2 consisted of 11 711 unique entries from 8 883 entities or users. Dataset 3 consisted of 3 556 unique entries from 1 938 entities or users. The overall metrics were calculated, and the key properties of the networks are summarized in Table 1.

Table 1. Graph metrics for the datasets

Graph Metric	Dataset 1	Dataset 2	Dataset 3
Vertices	4 935	8 883	1 938
Unique edges	5 753	9 369	2 131
Edges with duplicates	3 778	2 342	1 425
Total edges	9 531	11 711	3 556
Self-loops	2 524	3 032	641
Connected components	457	1 957	151
Single-vertex connected components	333	1 454	111
Graph density	0.000465	0.000200	0.001175

4 Results

Top vertices can be determined by the degree measure, or the betweenness centrality measure. Betweenness centrality would indicate more central, and arguably, the more influential entities, and can assist in determining the key influencers in the network. Few differences were found between the two measures, and the top vertices ranked by betweenness centrality is depicted in Table 2 below.

Table 2. Top vertices for the different datasets ranked by betweenness centrality

Dataset 1	Dataset 2	Dataset 3
IECSouthAfrica	AdvBarryRoux	PresidencyZA
News24	IECSouthAfrica	CyrilRamaphosa
Our_DA	Mandzenga	KhuselaS
SABCNewsOnline	GovernmentZA	GovernmentZA
SimonPGrindrod	tumisole	BiyelaSthe
SizweMpofuWalsh	News24	DIRCO_ZA
aneebh	ThabisoSithole	HloniNyetanyane
ali_naka	MsezaneSifiso	SANDFCorpEvents
AfricaInsights	SABCNewsOnline	TrafficSA

The next aspect we explored is the top hashtags per group, and for dataset 1 the top 5 hashtags per group are listed in Table 3.

Table 3. Top hashtags by frequency of mention for the largest groups in dataset 1

Top hashtags G1	Top hashtags G2	Top hashtags G3	Top hashtags G4
XseDay	SAElections2019	SAElections2019	SAElections2019
SAElections2019	FreedomDay	EFFMayDayRally	Elections2019
8-May	VoteDA	SMWX	ANC
FreedomDay	ANC	TshelaThupa	Elections
SADecides2019	DA	SouthAfrica	ACDPGAUTENG2019

The NodeXL visualization of the network graph for Dataset 1 is depicted in Fig. 1.

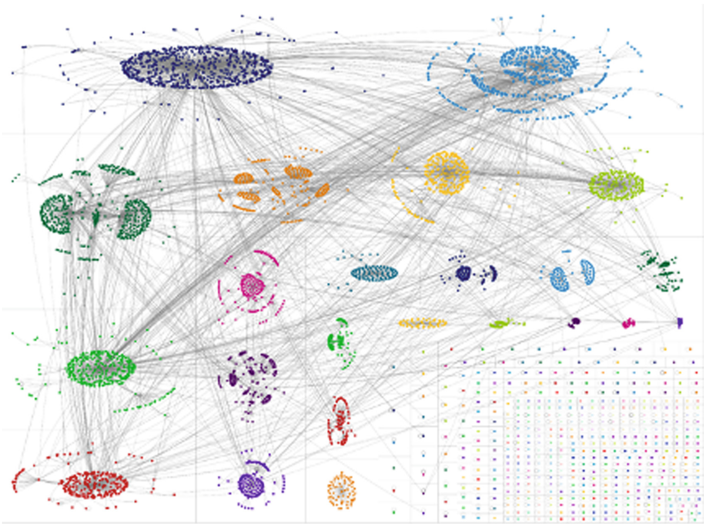

Fig. 1. Network visualisation for dataset 1: last week of campaigning (23–24 April 2019)

The top 5 hashtags per group for dataset 2 are listed in Table 4.

Table 4. Top hashtags by frequency of mention for the largest groups in dataset 2

Top hashtags G1	Top hashtags G2	Top hashtags G3	Top hashtags G4
SAElections2019	SAElections2019	8-May	IAmVotingEFF
XseDay	SADecides2019	XseDay	SAElections2019
SADecides2019	KnowYourDA	SAElections2019	Voting
IAmVotingEFF	OneSAforAll	SADecides2019	SADecides2019
Votingday	sabcnews	Voting	XseDay

The NodeXL visualization of the network graph for Dataset 2 is depicted in Fig. 2 and the network graph for Dataset 3 is depicted in Fig. 3 (Table 5).

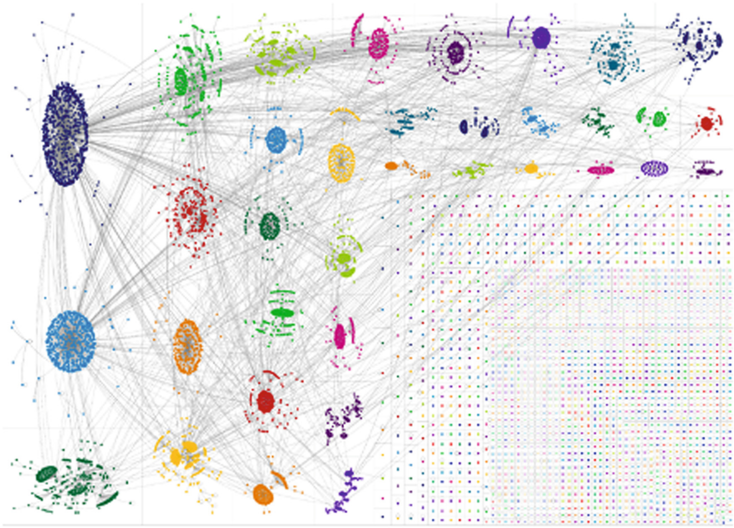

Fig. 2. Network visualisation for dataset 2: election day (7–8 May 2019)

The top hashtags in the five largest groups in all the datasets overlap significantly, and related closely to the top hashtags in the overall networks. It is also possible to determine that some of the smaller groups is centred around a specific political party and therefore referenced that party more prominently. In all the networks the connections within and between groups differs and therefore prominent clusters or groups are distinguishable. A much higher percentage of shared connections are within groups with few connections between groups.

Table 5. Top hashtags by frequency of mention for the largest groups in dataset 3.

Top hashtags G1	Top hashtags G2	Top hashtags G3	Top hashtags G4
PeoplesInauguration	PeoplesInauguration	PeoplesInauguration	SAInauguration19
eSwatini	NewDawn	SAInauguration19	PeoplesInauguration
SAInauguration19	SAInauguration19	NewDawn	NewDawn
WeThePeople	SABCNews	Democracy25	WeThePeople
CyrilRamaphosa	LoftusStadium	WeThePeople	SA

In all the networks, the Broadcast Network pattern is distinguishable, even though there are some similarities to the Community Cluster and Support Network patterns. The distinct hub-and-spoke groups in the graphs, particularly in Fig. 3, differentiates the Broadcast pattern from the Community Cluster, and the fact that we have several such hub-and-spoke groups distinguishes the network from the Support Network. The number of isolates in all the networks also supports a Broadcast Network and we therefore classified the networks therefore as Broadcast Networks. The Broadcast network pattern typically reflects Twitter traffic around breaking news stories or events with many people retweeting posts, or influencers and trend-setters whose tweets are retweeted often [19]. In a Community Cluster there are interactions between members of a group, which does not reflect the hub-and-spoke cores, and the only network that depicts some community clusters are in dataset 2 in Fig. 2 where some of the prominent smaller groups form interactive clusters. Such a cluster would be indicative of discussion between members of a group and not just repeated tweets.

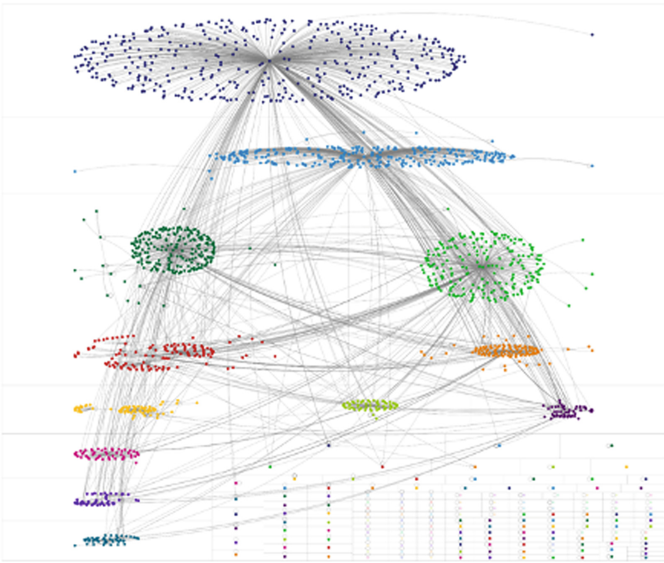

Fig. 3. Network visualisation for dataset 3: inauguration day 23–24 May 2019

5 Findings

All three of the datasets that were analysed can be classified as Broadcast Networks with a number of distinct groups. If we analyse the top hashtags it is clear that these groups were representative of the main political parties as well as the main media accounts that created traffic around breaking news and events. This is very prominent in Fig. 3, inauguration day, where very strong hub-and-spoke groups are discernible.

The social network analysis study done in previous research specifically in the USA found that political communities on Twitter form polarized crowds [11] where, even though both groups are focused on the same topic, there is little conversation between the groups. This study was, however, done in the United States of America where there are only two main political parties feature. Since South Africa has multiple political parties, multiple groups are formed, both on the general subject of the elections, as well as centred around specific political groups or influencers. This is evident in the results of all the datasets where the largest group is centred around the general topic – mostly using hashtags associated with election day, and most of the other groups mention different political parties or specific influential persons. In dataset 3 the groups are not focused on specific political parties but remain segregated. The percentage of connections shared among clusters is also noticeably higher than in the first two datasets. This could be attributed to the fact that the election results were already revealed, and the conversations focused more on the president than on the different political parties that were involved. This is also evident from the hashtags used within the groups – a much larger overlap is seen among groups.

Overall it is surprising that very little discussions are evident from our datasets. The general impression of Twitter users during this time was that there were many discussions and sometimes aggressive differences between individuals, however, when the networks are analysed, it is clear that most interactions were broadcasts from central vertices that represented political role-players.

6 Conclusion

Social media has become the prominent means of communication in recent decades, and conversations on social media create networks with identifiable groups. In Twitter specifically, connections as users reply to and mention one another in conversations form clusters or communities that provide insight. Depending on the subject in question, the groups and conversational structures differ. Mapping social networks can assist in understanding the different ways that individuals form communities and organize online. This information can assist individuals and organisations to make informed decisions on the ways that groups and conversations are formed online. Previous research found that six distinct patterns can be seen in the social structures and conversations of the users, depending on the topic in question [19]. This study investigated the network structures that are formed within the South African Twitter community around the 2019 elections. The NodeXL toolset was used for the social network analysis.

Three datasets were gathered based on significant periods and search criteria. The networks found in all three of the datasets were classified as Broadcast Networks where a large number of disconnected entities contribute to the network by broadcasting key topics and groups do not exhibit strong links to other entities and groups within the network. The network structures are evident when there are central or influential users, typically media houses and political parties. From our datasets very little discussions are evident within groups and most interactions were broadcasts from central vertices that represented political role-players even though the general impression of Twitter users during this time was that there are many more discussions between groups.

References

1. Fu, X., Jar-Der, L., Boos, M.: Social network analysis: interdisciplinary approaches and case studies. CRC Press (2017). https://doi.org/10.1201/9781315369594
2. Scott, J., Carrington, P.J.: The SAGE Handbook of Social Network Analysis. SAGE, London (2011)
3. Hanna, R., Rohm, A., Crittenden, V.L.: We're all connected: the power of the social media ecosystem. Bus. Horiz. **54**, 265–273 (2011). https://doi.org/10.1016/j.bushor.2011.01.007
4. Perrin, A.: Social media usage: 2005–2015. Pew Research Center (2015). (https://www.secretintelligenceservice.org/wp-content/uploads/2016/02/PI_2015-10-08_Social-Networking-Usage-2005-2015_FINAL.pdf)
5. Kaplan, A.M., Haenlein, M.: Users of the world, unite! the challenges and opportunities of social media. Bus. Horiz. **53**, 59–68 (2010). https://doi.org/10.1016/j.bushor.2009.09.003
6. Schivinski, B., Dabrowski, D.: The effect of social media communication on consumer perceptions of brands. J. Mark. Commun. **22**, 189–214 (2016). https://doi.org/10.1080/13527266.2013.871323
7. Wellman, B. (ed.): Networks in the Global Village: Life in Contemporary Communities. Westview Press, Boulder, Colo (1999)
8. Khan, G.F.: Seven Layers of Social Media Analytics: Mining Business Insights from Social Media Text, Actions, Networks, Hyperlinks, Apps, Search Engine, and Location Data. CreateSpace Independent Publishing Platform, Leipzig (2015)
9. Tichy, N.M.: Social network analysis for organizations. Acad. Manage. Rev. **4**, 507–519 (1979)
10. Kietzmann, J.H., Hermkens, K., McCarthy, I.P., Silvestre, B.S.: Social media? get serious! understanding the functional building blocks of social media. Bus. Horiz. **54**, 241–251 (2011). https://doi.org/10.1016/j.bushor.2011.01.005
11. Adamic, L.A., Glance, N.: The political blogosphere and the 2004 U.S. election: divided they blog 8 (2005)
12. Bennett, W.L.: The personalization of politics: political identity, social media, and changing patterns of participation. Ann. Am. Acad. Polit. Soc. Sci. **644**, 20–39 (2012). https://doi.org/10.1177/0002716212451428
13. Trottier, D.: Social Media, Politics and the State: Protests, Revolutions, Riots, Crime and Policing in the Age of Facebook, Twitter and YouTube. Routledge, Abingdon (2014). https://doi.org/10.4324/9781315764832
14. Hansen, et al.: - Do You Know the Way to SNA A Process Model for A.pdf
15. Rainie, H., Wellman, B.: Networked: the New Social Operating System. MIT Press, Cambridge, Mass (2012)

16. Rodrigues, E.M., Milic-Frayling, N., Smith, M., Shneiderman, B., Hansen, D.: Group-in-a-box layout for multi-faceted analysis of communities. In: 2011 IEEE Third Int'l Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third Int'l Conference on Social Computing, pp. 354–361. IEEE, Boston, MA, USA (2011). https://doi.org/10.1109/PASSAT/SocialCom.2011.139
17. Ma, D.: Visualization of Social Media Data: Mapping Changing Social Networks (2012)
18. Hansen, D.L., Rotman, D., Bonsignore, E., Rodrigues, E.M., Smith, M., Shneiderman, B., Capone, T.: Do you know the way to SNA?: a process model for analyzing and visualizing social media data. In: International Conference on Social Informatics, p. 10. IEEE (2012)
19. Social Media Research Foundation. https://www.smrfoundation.org/. Accessed 30 June 2020
20. Fay, D.: NodeXL: Network Overview, Discovery and Exploration in Excel. https://www.microsoft.com/en-us/research/project/nodexl-network-overview-discovery-and-exploration-in-excel/. Accessed 30 June 2020
21. Smith, M.A., Rainie, L., Shneiderman, B., Himelboim, I.: Mapping twitter topic networks from polarized crowds to community clusters. Pew Res. Cent. **20**, 1–56 (2014)
22. Hansen, D.L., Schneiderman, B., Smith, M.A.: Analyzing Social Media Networks with NodeXL: Insights from a Connected World. Morgan Kaufmann, Burlington, MA (2011)
23. Twitter. https://twitter.com/home. Accessed 20 June 2020
24. Clauset, A., Newman, M.E.J., Moore, C.: Finding community structure in very large networks. Phys. Rev. E. **70**, (2004). https://doi.org/10.1103/PhysRevE.70.066111