Analytical Tools

Semi-Public Space Conflicts and Alliances in Primary Metropolitan Centres: Sylvia Park, Mt Wellington, Auckland

LITERATURE REVIEW ON DIGITAL SPACE ANALYSIS METHODOLOGY

Angelo Bueno, Tanyalak Chalermtip, and Manfredo Manfredini

Auckland, 28/06/2019

[Working document GUS/SP1.3]

Contents

1. Network Analysis

1.1 Multivariate Networks: Unexpected Patterns and Relationships

Traditionally, interviews and questionnaires have been the preferred data source for recording complex social interactions due to their ability to capture emotive cues in written or verbal responses (Palla, Pollner, & Vicsek, 2009, p. 11). However, the studies that use them tend to focus on small and selective groups because of the great deal of time and manual processing (Palla et al., 2009, p. 11). This has often resulted in studies having a limited point of view, as a few individuals cannot accurately represent a wide demographic. This is particularly an issue when researching urban areas that relate to complex and diverse population groups.

Furthermore, when questionnaires are used, they require lengthy planning and testing (including addressing important privacy issues) and often collect written and verbal data with little respondent engagement and focus.

In contrast, participatory sensing systems have enabled researchers to involve public participation (Kuznetsov & Paulos, 2010, p. 21), by giving them access to visual-locative and multivariate data. Due to the ubiquity of smartphones, these have become an essential source of information for studying urban issues. Recent studies have increasingly utilised these technologies to investigate how different groups of people use Instagram to represent important locations within the city (Boy & Uitermark, 2017, p. 612). The data can be intuitively visualised and explored using a range of tools which allow researchers to manually select and isolate areas to detect clusters that hold similar attributes, or to detect, locate, and highlight, etc. highly correlated attributes across an entire network.

John D. Boy and Justus Uitermark have pointed out that interactions which occur naturally on Instagram circumvent these issues as it encourages interaction (through likes, comments, replies, follows), is highly visual (containing videos, images, emoticons), and content can be uploaded or accessed at any location and time via the mobile application (Boy & Uitermark, 2016, p. 2). In the study, "Reassembling the City through Instagram," an initial description of the "functionality and conventions" on Instagram was from the perspective of 16 active users between the ages of 19–51. Many of the users had identical or similar occupations (Boy & Uitermark, 2017, p. 614). As the sample was relatively small and homogenous, it was a minor component of the research.

However, the major focus of the investigation was on 400,000 Instagram posts within Amsterdam, about which they obtained information through a custom scraper they developed called *Kijkeens* (Boy & Uitermark, 2018). Unlike the results from the interviews, the scraped data could be concisely visualised in a series of graphs. A heatmap was produced to show the physical distribution of posts that were tagged to different points of interest throughout Amsterdam (Boy & Uitermark, 2017, p. 614); the proportion of highly popular users was shown on an in-degree distribution graph (Boy & Uitermark, 2017, p. 618); and the strength of interactions between clusters of users was simplified by merging nodes and using edge weights for easier readability (Boy & Uitermark, 2017, p. 620). Andreas Kerren et al. have characterised different visualisation approaches for multivariate datasets such as *pixel-based approaches* (e.g., scatterplots), *projection methods* (e.g., multi-dimensional scaling), *axis-based approaches* (e.g., parallel coordinate plots), and *icon-based approaches* (e.g., Chernoff faces) (Kerren, Purchase, & Ward, 2014, p. 3).

Furthermore, van den Elzen and van Wijk (2014) developed visualisation as an exploratory process to discover relationships in multivariate data in addition to presenting end results. As such, various tools have been developed to find outliers, patterns and trends within multivariate networks. *DOSA* is a tool made up of four interactive panels that allow data to be selected and analysed in different levels of detail. The *filter panel* has a set of controls which are used to hide nodes and edges which have attributes that fall outside of the range specified by the user. The *detail panel* allows further refinement by manually dragging multiple selection boxes; the s*election panel* toggles between the visibility of each selection box; and the *overview panel* automatically produces an "infographic-style" diagram based on the other three panels (van den Elzen & van Wijk, 2014, pp. 2310–2319). Providing an intuitive and dynamic means of exploration allows different levels of focus such as within a selection, between selections, on the background influences of a selection, or the combination of all these approaches (van den Elzen & van Wijk, 2014, pp. 2310–2319).

Similarly, "GraphDice" (Bezerianos, Chevalier, Dragicevic, Elmqvist, & Fekete, 2010) also explores multivariate networks with two major differences from *DOSA*. It can view plot graphs of every possible attribute combination in an overview matrix allowing the user to immediately notice highly correlated attributes without the need to check every combination individually. Secondly, a lasso selection is used instead of a box selection, and nodes can be added or subtracted from the selection (Bezerianos et al., 2010, pp. 866–867). In addition to speeding up workflow, these features give full control for defining areas of the network for queries.

Adopting social media as a data source for investigating urban public issues creates new research opportunities due to its participatory, visual, and accessible nature. By tracking the number of likes and comments, denser clusters in the network can be detected and compared. The visualisation and intuitive handling of information allows unexpected relationships to be discovered. As Instagram content can be uploaded by anyone, an extensive data source has been built which does not under-represent or discriminate against any demographic groups. Overall, using social media for research purposes, enables researchers to highlight how effects are manifested across an entire system rather than just individuals. This is particularly suited to understanding how groups with distinct attributes respond differently to public space.

1.2 Network Partitions: Group Characteristics and Interactions

The pure number of nodes and edges does not describe complex interactions, therefore studies on entire social networks tend to focus on topology (Smith, Rainie, Shneiderman, & Himelboim, 2010). The study on individual nodes does not uncover meaningful relationships and does not show the role of centrality in influencing the behaviour of entire specific groups. However, by partitioning the network into exclusive groups, researchers can investigate the level of coherence within groups, and between groups, and the effect this has on their interactions and points of interest.

Communities based on interactions can be structured differently using various detection methods to meet specific research objectives; they are also denser than clusters formed by a set of shared attributes. One method results in hierarchical clusters which assign nodes into groups that are embedded inside of a series of larger groups. The nodes simultaneously belong to many groups but become more loosely connected to members further along the chain (Soler, Tencé, Gaubert, & Buche, 2013, p. 492). Detection methods which form exclusive groups do not have a tiered structure as they assign nodes to only one group (Glenn, 2001, p. 6). As this ensures the maximum differences between groups and the highest similarity within groups, it is particularly suited for comparative analysis.

On the other hand, the multi-resolution approach of the Louvain method repeats several passes of merging partitions (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008, p. 3). This process results in dense clusters that are distinct but can also be zoomed in or out until a desired resolution is meet (Blondel et al., 2008, pp. 5–8). Adjusting the resolution allows researchers to find an appropriate balance between the size and number of communities. While the first step of this method is to make every node its own partition, to detect the many smaller communities that have a less impact, it is essential to identify the number of communities that will then be used for an in-depth analysis, which retains as much information as possible. An example of the application of this process is Boy and Uitermark’s (2005) study on a group of 300,000+ Instagram users, where eight clusters were created by filtering out communities with fewer than 100 people (Boy & Uitermark, 2005, p. 2). As the filter can be set to any arbitrary value, it may be replaced with a value which is determined by a percentage of all nodes in the network. This approach would create a filter which adapts network size to retain more information. Additionally, consistently sized communities ensure better comparisons between central nodes, such as community leaders or parochial places (Boy & Uitermark, 2016, pp. 5–6, 12). Two communities with an extreme size difference are unsuitable for this comparison, as central elements in small groups are volatile and can change with just a few interactions, while larger groups need a stronger effect to shift centrality rank. Once the partitions are deemed suitable for comparison, the separation between groups can be measured by using the index of dissimilarity (Boy & Uitermark, 2016, p. 5). Conversely, the interaction between groups may be measured by taking the mean average of exchanges for all members of both communities (Boy & Uitermark, 2016, p. 6).

It is often theorised that people in close contact, such as those within the same community, share many of their attributes (De Nooy, Mrvar, & Batagelj, 2018, p. 59). Indeed, in “How to Study the City on Instagram” each community was similar enough to be characterised by a single label such as “Lifestyle Vanguards” or

“Cultural Entrepreneurs” (Boy & Uitermark, 2016, pp. 7, 10). Matching these labels with the levels of interaction between groups has revealed how these two distinct communities are in fact highly connected. Further investigation may explore whether these subsets are in fact interdependent. Dror Kenett et al. have pointed out that links between two separate networks may be vital for both to operate properly, such as in communication networks which rely on power grids for energy while power grids also need communication networks for control information (Kenett, Perc, & Boccaletti, 2015, p. 2). Rather than focusing on coherent groups, this approach explores unexpected relationships that form when seemingly opposite groups are highly connected. This is particularly important for understanding how people share a public space or communicate on the same social media platform while holding fundamental differences such as attitudes or beliefs, status or social class, workplaces or educational institutions, and informal roles or occupational positions (McPherson, Smith-Lovin, & Cook, 2001, pp. 415–437).

1.3 Real-World Networks: Properties and Mechanisms

Centrality considers the direction and number of interactions to find nodes that hold important roles. While degree centrality provides a general level of interaction by counting direct edges of a node (Hanneman & Riddle, 2011, p. 360), a wide variety of measures are used for more specific purposes to indicate such things as: authoritative nodes that receive considerable attention (in-degree centrality) (De Nooy et al., 2018, p. 189); nodes which send out many messages – out-degree centrality (Wasserman & Faust, 1994, p. 6); nodes that can easily reach all parts of the network – closeness centrality (Wasserman & Faust, 1994, p. 6); or nodes that bridge different regions of a network – betweenness centrality (Wasserman & Faust, 1994, p. 6).

Albert-László Barabási has identified growth and preferential attachment as the primary causes for edges to form in a manner which leads to certain nodes taking on important roles in the network such as hubs (Barabási, 2016, p. 6). Barabási states that as nodes are introduced into the network, they tend to interact with already well-connected nodes and the accumulative effect leads to extreme differences in in-degree centrality (McPherson et al., 2001). While social media has made it possible to connect a wider number of people, it has been noted that these connections reflect relationships that have been established in real life (McPherson et al., 2001).

These mechanisms are important to studying real-world networks as they have a degree distribution and centrality measures that cannot be reproduced within random graphs. The edges of random graphs are generated by a process which uses a probability value to determine the chance that a connection is formed between any two nodes (Newman & Park, 2003, p. 2). This parameter produces a normal distribution with a well-defined average calculated on the unevenness of real-world networks (Barabási & Bonabeau, 2003, p. 63), allowing researchers to define features such as hubs or bridges (Wasserman & Faust, 1994, p. 6). In contrast, real social networks have been known to exhibit a power-law degree distribution which is characterised by a few highly connected nodes and a high clustering of less connected nodes (Barabási, 2016, p. 9). This leads to two major research directions in complex networks; one is to understand the causes of extreme differences of degree distribution in scale-free networks (Barabási & Bonabeau, 2003, p. 63); and the other investigates the small-world phenomena where nodes are not neighbours of every other node but can easily reach all part of the network within just a few hops (Watts & Strogatz, 1998, p. 440).

1.4 Two-Mode Networks: Projection

“Modes” relate to the number of objects or ideas which a node represents (Borgatti, Everett, & Johnson, 2018, p. 267). As one node type is sufficient for most research purposes, most metrics have been developed for one-mode networks (Latapy, Magnien, & Del Vecchio, 2008, p. 31). By the same logic, two-mode networks introduce a second node type forming two distinct sets of nodes (Latapy et al., 2008, p. 31). This is best suited for visually understanding collaborative networks as it may differentiate people from their associated organisation through the use of colour (Wasserman & Faust, 1994, p. 103). This has benefited the study of a wide range of complex networks related to technology (Internet); social (collaboration networks), and linguistic ones (co-occurrence networks) (Latapy et al., 2008, p. 33). While it is possible to introduce an unlimited number of node types, the calculation of standard network metrics has become increasingly problematic (Wasserman & Faust, 1994, p. 35).

Experimental methods for two-mode networks have been developed but not adopted within mainstream network analysis (Latapy et al., 2008, p. 31). Instead, standard practice has been to transform two-mode networks via projection to allow conventional methods to be used. This process forms a link between any two people who hold membership to the same node (Zhou, Ren, Medo, & Zhang, 2007, p. 1). This connects affiliated nodes but at the cost of losing proper edge weights (Zhou et al., 2007, p. 1).

In the case of Instagram, the dataset bypasses these projection issues by avoiding the process completely. A two-mode network consisting of Instagram posts and people can be transformed into a one-mode network by replacing the post ID with its author. These two versions of the network can be analysed separately to benefit from the accurate visualisations of the two-mode network, while using the one-mode network to preserve edge weights and calculate accurate centrality values.

2. Semantic Analysis

Content analysis is the "summarizing, quantitative analysis of messages that relies on the scientific method (including attention to objectivity, intersubjectivity, a priori design, reliability, validity, generalizability, replicability, and hypothesis testing) and is not limited as to the types of variables that may be measured or the context in which the messages are created or presented" (Neuendorf, 2002, p. 10). Written text or messages are a form of communication and expression which reveal people's thoughts, beliefs, feelings, and personalities. Significantly, social media messages allow users to communicate and express their opinions through different textual elements such as status, captions, comments, and hashtags. These short messages and keywords contain people's perception of various topics in their everyday lives. Thus, researchers in different fields utilise these semantic contents to investigate people's psychological and behavioural factors to understand or predict a probability of a person's or community's communicative act (Schwartz & Ungar, 2015, p. 85).

2.1 Content Analysis Types and Techniques Used in Semantic Fields

Two major types of automatic content analysis techniques, *closed vocabulary* and *open vocabulary*, different approaches and a variety of methods are used to examine messages. On the one hand, closed vocabulary techniques involve the definition of a given set of rules prior to analysing messages. The most common methods to identify and count the usage of words contained in messages use either *manually compiled* or *crowd-sourced dictionaries*. The manual dictionaries are composed of pre-defined categories of words (Schwartz et al., 2013). The crowd-sourced dictionaries are composed of categories of pre-selected words created with the identification of words' association strength through score-based rating methods (Schwartz & Ungar, 2015, p. 81).

On the other hand, open vocabulary techniques analyse messages relying on dynamic lists of words, which derive their dictionaries from texts, and topics. Deriving dictionaries from text begins with a *tokenisation* process, in which messages are split into individual words, and words associated in categories that are generated automatically based on their overall correlation (Schwartz & Ungar, 2015, p. 82). Topics are also generated automatically by grouping words that often co-occur in the same cluster in large categories (Schwartz & Ungar, 2015, p. 83).

Closed and open vocabulary techniques offer various methods of analysing messages, providing different outcomes for different analysis purposes. The closed vocabulary technique enables researchers to isolate a set of words to focus on a specific topic. Open vocabulary techniques are significant when analysing incoherent sets of short messages, such as the social media ones, because they can detect unpredictable words belonging to unexpected languages and slang as well as abbreviations. The outcomes of this type of analysis produce complex information that requires further processing (e.g., results often include three-digit numbers of topics) (Schwartz & Ungar, 2015, p. 83).

These two types of content analysis are often used in combination, enabling a focus on a restricted number of significant topics with the use of context-specific closed vocabularies that are developed through open vocabulary analyses.

2.2 Semantic Analysis-Related Work

Content analysis enables researchers to investigate semantic fields on social media to understand the users' behaviour and observe users' perspectives. For example, a research paper by Lydia Manikonda, Yuheng Hu, and Subbarao Kambhampati examines the average wordcount and average characters used on Instagram comments to study users' commenting behaviour on Instagram posts (Manikonda, Hu, & Kambhampati, 2014). Moreover, the researchers discover Instagram popular hashtags based on their datasets, enabling them to identify groups of people with similar interests (Manikonda et al., 2014).

Similarly, Jin Yea Jang et al. study the variation between teens' and adults' engagement on social media by investigating a number of photos, likes, hashtags, comments, followers, and followings of the two age groups (Jang Han, Shih, & Lee, 2015). They also investigate teens' and adults' interests based on their hashtags by categorising the hashtag topics (for example, arts/photo/design, entertainment, fashion/beauty, food, nature, sports/wellness, etc.) following two popular hashtag websites–tagsforlikes.com and tagstagram.com–and calculate the ratio of adults and teens associated with each hashtag topic to identify the main interests of the two age groups (Jang et al. 2015, p. 4041).

Data-processing techniques can also include basic cleaning and normalisation processes such as the one used by Basit Shahzad et al. (2017). To discover and classify users' interest topics in Twitter messages (tweets), they pre-process collected data to remove stop words, such as *a, an, the;* normalise recurrent chat terms, such as "f9," "lmao," "cuz," converting them into standard words with Chat Words Dictionary and PyEnchant library; and tag the term types using part of speech tagging (POS) to identify and collect only the nouns (Shahzad et al. 2017, pp. 132-134).

Tweets are then labelled and attributed to relevant topics (allowing for multiple labelling and attribution of individual tweets). Their classification of the topics includes 30 categories that further articulates the 18 general trending topics set formulated by Lee et al. (Shahzad et al., 2017, pp. 133). It also identifies the topics' perception (as positive or negative) and popularity (showing the relative of the popularity of the main topics) (Shahzad et al., 2017, pp. 133–135).

While these studies utilise semantic elements on social media to understand people's behaviour and identify their interests, other researchers explore semantic fields on social media to observe people's perspective on public events and places. For example, Shuhua Liu and Patrick Jansson use keywords to discover city events on social media to explore clusters of events and understand the city dynamic. They collect Instagram data from the Helsinki metropolitan region, which detects 47 different languages in the comments and posts (Liu & Jansson, 2017, p. 2). The datasets are filtered to remove non-essential data and only focus on English to ensure consistency. The events-detection method consists of three major steps: firstly, the researchers define targeted event types (urban festival, sports, safety, security) and initial keywords (festival, sport, safety, security) (Liu & Jansson, 2017, p. 3). Secondly, they expand the initial keywords through semantic similarity analysis using Word2vec software to collect more related terms. Lastly, they only extract the relevant Instagram data based on the expanded terms from the entire dataset to visualise and analyse the results (Liu & Jansson, 2017, p. 4). The expansion of the keywords allows the researchers to find more targeted events on social media and discover significant festivals occurring around the city.

Furthermore, social media semantic analysis has been used to observe public opinions on a place; this technique is what the researchers, Linlin You and Bige Tunçer, call "geo-sentiment analysis," and can be aimed at providing a better understanding of spatial perception to orient strategies and action in urban planning and governance (You & Tunçer, 2016, p. 693). The "crowd-calibrated geo-sentiment analysis mechanism" (CGSA) developed by You and Tunçer utilises the location-tag function on social media to define specific coordinates and retrieve semantic data from precisely designated areas only (You & Tunçer, 2016, p. 695). The data are then analysed by using proprietary tools that categorise expressions into three groups – positive, neutral, and negative (You & Tunçer, 2016, pp. 696–697). The results are processed through clustering analysis and time-series analysis to identify articulations and trends in public opinion on specific locations and evaluate the effect of spatial planning and management (You & Tunçer, 2016, pp. 699).

In tourism studies, semantic analysis uses information from social media (i.e. content, geo-location tags and metadata) to examine visitors' perspectives of a place. Yuri Rykov et al. investigated Instagram data to assess the variation of people's experiences in a space. A dataset of images, dates and times, location coordinates, user IDs, and hashtags are extracted from St. Petersburg (Rykov, Nagornyy, & Koltsova, 2016, p. 110). The images are analysed firstly using Google Cloud Vision API service to create labels of recognised image content, then with Cortext Manager software to group the labels into different clusters (Rykov et al., 2016, p. 110). Finally, the image coordinates allow each cluster to be visualised on a geographical map using QGIS software (Rykov et al., 2016, p. 111). The result enables identification of the primary semantic domains generated by social media users, as well as visualising the relevant elements of the image of a given urban space (Rykov et al., 2016, p. 111).

Additionally, semantic analysis has also been used to investigate visitors' experiences, discover the places they visited to understand visitors' perceptions and identify the destination image. Linda S. Lai and Wai Ming's study on Macao used data (comments and reviews) from tourism-related websites (e.g., Travelblog.org, Travelpo.com, and Tripadvisor.com) (Lai & To, 2015, p. 144). They processed semantic data using lexical and statistical analysis tools to transform the words into a similar format such as consistent spelling or one-word format (Lai & To, 2015, p. 145). The words are then analysed, using a lexical analytic tool called WordSmith, to identify the keywords, keyness values of the words, and the frequency words appeared on the dataset (Lai & To, 2015, p. 143). These keywords are identified into clusters of related words in the factor-analysis phase, and the lexical mapping software, Leximancer, categorises the clusters of keywords into different themes and visualises them as a network map (Lai & To, 2015, p. 144). The result of the keyword themes enables the researcher to identify the majority destination image of Macao as a casino city with cultural heritage as its background (Lai & To, 2015, p. 147).

Significantly, the researches utilise semantic analysis to explore people's behaviour, interests, and perceptions, especially Instagram semantic content, which is commonly used along with geo-location tags to conduct research on specific locations. Thus, this allows an investigation of public opinion on public spaces in specific locations. Moreover, there is a bridge between image and semantic analysis as shown, in Yuri Rykov's research, by using Google Cloud Vision API service to identify word labels of recognised objects within image contents. The automation software is also widely used in semantic analysis to process a large number of data and minimise human errors. However, since the analysis relies on software with limited accuracy, You and Tunçer have mitigated this issue with the creation of a toolset that combines different software.

Research aimed at improving public space usage to improve people’s wellbeing, such as Give Us Space, can adopt an integrated longitudinal network and semantic analysis on social media data to better understand socio-spatial relationality and detect trends in situated users’ activities, providing significant new insight into the understanding of emerging problems of semi-public space.

References

Barabási, A. -L. (2016). *Network science*. Cambridge, England: Cambridge University Press.

Barabási, A. -L., & Bonabeau, E. (2003). Scale-free networks. *Scientific American, 288*, 60–69.

Bezerianos, A., Chevalier, F., Dragicevic, P., Elmqvist, N., & Fekete, J. -D. (2010). Graphdice: A system for exploring multivariate social networks. *Computer Graphics Forum*, *29*(3), 863–872.

Blondel, V. D., Guillaume, J. -L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(10), 10008.

Borgatti, S. P., Everett, M. G., & Johnson, J. C. (2018). *Analyzing social networks* (2nd ed.). Bristol, England: Sage.

Boy, J. D. *Kijkeens: A tool for researchers*. Retrieved from https://zenodo.org/record/50497#.W7ZC0Wgza70

Boy, J. D., & Uitermark, J. (2005). *Capture and share the city: Mapping Instagram’s uneven geography in Amsterdam.* Paper presented at RC21 International Conference on The Ideal City: Between Myth and Reality. Representations, Policies, Contradictions and Challenges for Tomorrow's Urban life, Urbino, Italy. https://www.rc21.org/en/wp-content/uploads/2014/12/B1-boy-uitermark-instagram-paper.pdf

Boy, J. D., & Uitermark, J. (2016). How to study the city on Instagram. *PloS ONE, 11*(6), e0158161.

Boy, J. D., & Uitermark, J. (2017). Reassembling the city through Instagram. *Transactions of the Institute of British Geographers, 42*, 612–624.

De Nooy, W., Mrvar, A., & Batagelj, V. (2018). *Exploratory social network analysis with Pajek.* Cambridge, England*:* Cambridge University Press.

Fung, G. (2001). A comprehensive overview of basic clustering algorithms. *CiteSeerX,* 1–47.

Hanneman, R. A., & Riddle, M. (2011). Concepts and measures for basic network analysis. In J. Scott & P. J. Carrington (Eds.), *The Sage handbook of social network analysis* (pp. 340–369). London, England: Sage.

Kenett, D. Y., Perc, M., & Boccaletti, S. (2015). Networks of networks–An introduction. *Chaos, Solitons & Fractals, 80*, 1–6.

Kerren, A., Purchase, H. C., & Ward, M. O. (2014). Introduction to multivariate network visualization. In A. Kerren, H. C. Purchase, & M. O. Ward (Eds.), *Multivariate network visualization* (pp. 1–9). Berlin, Germany: Springer.

Kuznetsov, S., & Paulos, E. (2010). Participatory sensing in public spaces: Activating urban surfaces with sensor probes. In *Proceedings of the 8th ACM Conference on Designing Interactive Systems*, 21–30.

Jang, J. Y., Han, K., Shih, P. C., & Lee, D. (2015). Generation like: Comparative characteristics in Instagram. In *CHI '15 Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 4039–4042.

Lai, L. S., & To, W. M. (2015). Content analysis of social media: A grounded theory approach. *Journal of Electronic Commerce Research, 16*(2), 138–152.

Latapy, M., Magnien, C., & Del Vecchio, N. (2008). Basic notions for the analysis of large two-mode networks. *Social Networks, 30*(1), 31–48.

Liu, S., & Jansson, P. (2017). City event identification from Instagram data using word embedding and topic model visualization. *Arcada Working Papers*, 1–8.

Manikonda, L., Hu, Y., & Kambhampati, S. (2014). Analyzing user activities, demographics, social network structure and user-generated content on Instagram. *CoRR abs*/1410.8099, 1–5.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology, 27(*1), 415–444.

Neuendorf, K. (2002). *The content analysis guidebook*. Thousand Oaks, CA: Sage.

Newman, M., & Park, J. (2003). Why social networks are different from other types of networks. *Physical Review, 68*(3 Pt 2), 036122.

Palla, G., Pollner, P., Barabási, A. -L., & Vicsek, T. (2009). Social group dynamics in networks. In T. Gross & H. Sayama (Eds.), *Adaptive networks* (pp. 11–38). Cambridge, MA: Springer.

Rykov, Y., Nagornyy, O., & Koltsova, O. (2016). Semantic and geospatial mapping of Instagram images in Saint-Petersburg. In *Proceeding of the AINL FRUCT 2016 Conference*, 110–113.

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal,… Ungar, L. H. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLoS ONE, 8*(9), 1–16.

Schwartz, H. A., & Ungar, L. H. (2015). Data-driven content analysis of social media: A systematic overview of automated methods. *The ANNALS of the American Academy of Political and Social Science*, *659*(1), 78–94.

Shahzad, B., Lali, I., Nawaz, M. S., Aslam, W., Mustafa, R., & Mashkoor, A. (2017). Discovery and classification of user interests on social media. *Information Discovery and Delivery, 45*(3), 130–138.

Smith, M. A., Rainie, L., Shneiderman, B., & Himelboim, I. (2014). Mapping Twitter topic networks: From polarized crowds to community clusters. *Pew Research Center, 20*, 1–56.

Soler, J., Tencé, F., Gaubert, L., & Buche, C. (2013). *Data clustering and similarity*. Paper presented at FLAIRS Conference.

van den Elzen, S., & van Wijk, J. J. (2014). Multivariate network exploration and presentation: From detail to overview via selections and aggregations. *IEEE Transactions on Visualization and Computer Graphics, 20*(12), 2310–2319.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications.* Cambridge, England: Cambridge University Press.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of ‘small-world’ networks. *Nature, 393*(6684), 440.

You, L., & Tunçer, B. (2016). Exploring public sentiments for livable places based on a crowd-calibrated sentiment analysis mechanism. In *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining,* 693–700.

Zhou, T., Ren, J., Medo, M., & Zhang, Y. -C. (2007). Bipartite network projection and personal recommendation. *Physical Review, E 76*(4), 046115.