Recommending Communities at the Regional & City Level

John P. Verostek

markITview Research
Cambridge, MA, USA
{john} @markitview.com

ABSTRACT

An exploratory study was undertaken to compare the community ecosystems of Boston, Silicon Valley, and New York City. The motivation was to understand what new groups could be *recommended* to improve the community landscapes. In particular, business and technology Meetup groups were studied with respect to type, number, size; and the interconnectedness across groups. The core data for the study was obtained via Meetup's API which enables access to community information. Using a trained, manually coded subset of titles and topic tags, machine learning was performed to categorize groups. Summary statistics were then produced and cities compared. For example, Boston's tech Meetup foundation was built upon programming languages such as PhP and Java. In contrast, New York City historically has cultivated a stronger business networking culture; however this type of group has been less pronounced in Boston. Next, social network analysis was utilized to identify key community groups via centrality measures and to identify "nearest neighbors" for certain groups. The content-based machine learning output was combined with the social network analysis to create a tiered, hybrid recommendation system. Lessons learned from the study would appear generalizable to cities beyond the three initial covered. For example, an economic advisor could recreate the community evolution of a city, and perform comparative regional analysis to then develop new recommendation for community growth. The next project steps were to move from 'proof of concept' towards developing a prototype community recommendation system.

Author Keywords
Meetup.com, community, recommendations, social network analysis, economic development, machine learning

ACM Classification Keywords
H.3.1 [Information Storage and Retrieval]: Content Analysis and Indexing.

INTRODUCTION

One research stream within economic development has been knowledge diffusion (or spillover) within a region. In 1994, Saxenian analyzed the electronic industries in Silicon Valley and Boston's Route 128, and found that informal networks helped Silicon Valley with economic growth [1]. Powell, et al, studied the biotechnology industry using network connectivity as a lens for which to understand knowledge transfer [2]. In general, these and other pre-Internet studies of knowledge flow relied upon qualitative research and quantitative measures such as scientific journal citations, patents, and company financials.

In the past decade, online social media, networks, and communities have drawn much more interest than their offline counterparts. Traditional face-to-face communities, whose history was illustrated in the Robert Putnam's book Bowling Alone [3], have received less attention from a business standpoint or academic research perspective. Furthermore, the line between offline and online has become blurred such as with Meetup.com straddling both worlds. A notable past study covering this platform was Weinberg and Williams analysis of Howard Dean's 2004 presidential run and his utilization of Meetup [4]. At that time Meetup.com was in its infancy having launched shortly after 9/11 in 2002.

More recently Meetup has seen accelerated growth especially in New York City (its home town), Silicon Valley, and also in London, Washington DC, Chicago, Boston, Toronto, Los Angeles among several others. Groups were started by organizers who each work with their members to hold events. Communities exist across a wide variety of interests often reflecting preexisting local activities. However, this study focuses on business and technology which has been a popular topic in Cambridge, MA (and in particular Kendall Square near MIT). Nonetheless, the findings appear generalizable across many types of groups. Additionally, Meetup cross-collaboration has occurred naturally among groups, i.e. co-hosting events or cross-promoting. However, a city or regional-level perspective was chosen for this research in order to understand broader community dynamics. After these ecosystems have been described, then recommendations could be made to create, connect, or expand existing groups (or clusters of groups) to positively impact a community.

Hence, today the evolution and acceleration of online social media platforms (such as LinkedIN and Meetup), have made visible previously unavailable data on communities. Regional community ecosystems can be more easily compared to understand similarities and differences.

The paper was organized as follows: first the primary research objectives were described; followed by a walkthrough of the data sources and necessary extensions to make the raw data useful for further analysis. Since this project has been a work-in-progress, preliminary results were then presented. This short paper concluded with near-term plans for next research and development steps.

PROPOSED RESEARCH APPROACH

The primary objective of this exploratory and descriptive research has been to develop a community recommendation system. The research was conducted across three inter-related dimensions which were represented as dyads:

- Community-Community: The goal was recommend to *existing* Meetup groups potential opportunities for collaboration with related local tech and business groups. This approach of finding similar communities was applied to groups in a home city; as well as other cities. An objective of these of recommendations was to *further integrate existing* communities. Groups share common members such that collaboration among groups may be beneficial such as with cross-pollination of ideas and knowledge. This objective was achieved primarily by understanding group-member overlap utilizing social network dynamics.

- City-City: The goal was to recommend *new* groups for a region. The recommendations were based on groups in other cities; but not present locally. For example, the Boston Predictive Analytics[1] was formed after discovery of NYC Predictive Analytics[2]. Similarly, Hacks & Hackers[3], a journalism meets software development group, has been adopted in several cities through the efforts of individuals who had learned of this group's presence on Meetup.

- Member-City: The goal was to recommend Meetup groups based upon aggregate interests of members of a local community. Please note that Meetup currently recommends *existing* groups to members based upon individual interests; therefore the focus here was to recommend *new* groups based on unmet interests.

Among the above three dyads (or levels of analysis) there existed a strong probability that the recommendations from each may conflict. For example, another city may have a strong industry presence that seldom exist in other places,

e.g. the uniqueness of the movie industry in Los Angeles. Conversely, a group like Hacks/Hackers given its journalism dimension would seem applicable to many cities. The above highlights that expansion of this project to include more city-specific attributes might improve the recommendations.

DATA SOURCES AND PREPARATION

Meetup's API includes several different methods that cover cities, groups, members, interests (or topics), events, and rsvp information. Foremost, Meetup.com API's was utilized to pull state and city information on community demographics for the United States.[4] Next, community group and member information including their associated "interests" were pulled for Silicon Valley, New York City, and Boston. These cities were chosen given their strong technology entrepreneurship leadership, high Meetup adoption, and as Cambridge, MA has been the home of the researcher for several years. A geographic factor by proxy of zip code was used to help identify regional boundaries.

A key to classifying the groups was utilizing topic tags; however this data exhibits a very long-tail such that it was necessary to create main categories and sub-categories in order to then enable regional comparisons and develop recommendations. Categories were created by manually coding over a thousand Meetup groups. Main categories created included technology and business dimensions; as well as travel, dining, music, outdoors, recreation, fitness, photography, singles, parenting, religion, et al.

Using machine learning classification techniques, thousands of groups from Boston, New York City, and Silicon Valley were then categorized across over thirty main categories. Specifically, R was used for data pulls, data management, and analysis. The library, RTextTools[5], was utilized for text processing and machine learning. This library contains well outlined steps for performing supervised learning. Multiple learning approaches were available including SVM, Maximum Entropy, Neural Networks, and others. Several iterations were performed towards identifying the best approach; as well as providing feedback for additional manual coding in order to improve model accuracy.

Subsequent analysis of the technology and business groups revealed patterns, particularly city differences, pertaining to the types of groups. Therefore, technology and business categories were further sub-segmented. For example, technology topics included software development, mobile, cloud, etc. Business categories covered professional networking, entrepreneurship, careers, jobs, marketing &

---

[1] http://www.meetup.com/Boston-Predictive-Analytics/

[2] http://www.meetup.com/NYC-Predictive-Analytics/

[3] http://meetupbos.hackshackers.com/

[4] Many thanks to Vipin Sachdeva of IBM Cambridge for helping with data pulls, and for showing me better ways of performing this step.

[5] http://www.rtexttools.com/

sales, and industry-specific verticals, such as healthcare and clean energy.

These new categories enabled recommendations to occur at multiple levels: category, sub-category, and specific group. For example, for a particular city a recommendation might be to create more "business" groups; or perhaps to develop a sub-category-level 'thematic' group such as "startups"; or be more specific, and recommend a community group pertaining to "healthcare information technology startups".

Meetup.com member profile also includes fields for complementary social media platforms including Twitter, LinkedIN, and Facebook. However, on Meetup these social media fields, as well as member interests; were optional fields thereby reducing the available sample.

With respect to the different levels of inquiry additional data processing was necessary towards setting the stage for further analysis:

- Community-Community: Group and member data was transformed into group-member dyads; also known as an edgelist. An R package, called 'tnet' developed by Tore Opsahl[6] was utilized to find edge weights as determined by member overlap between two groups. The output from this R library enabled bipartite social network analysis to quantify the overlap among groups (using group-member dyads). Gephi, a social network software package, was then utilized to perform analysis to derive network characteristics such as "centrality".[7] Gephi also was used to create data visualizations. Hierarchical cluster analysis was also performed to under the structure a region's ecosystem.

- City-City: Groups across cities were compared using text-based machine learning to ascertain the similarity or dissimilarity of groups. The text dataset consisted of group names with up to fifteen topic tags. Where applicable, the social graphs of city-groups were also compared to make more specific recommendations.

- Member-City: Meetup member interests were pulled by using their API; as well as extracting the interests specified by organizers for their groups. Meetup members were not limited to a set number of topics such that some had listed dozens upon dozens of interests. A common Meetup tag language simplified the analysis between members and groups. LinkedIN groups and skills were found often times to be visible on profiles. Please note that the initial exploratory research for LinkedIN was a subset of its members; specifically people who were also on Meetup's platform.

PRELIMINARY RESULTS AND DISCUSSION

Longitudinal analysis on community formation was also performed and showed that over the past few years Meetup.com has seen accelerated growth; especially for technology and business groups. Analysis at the category level showed Silicon Valley to have the highest percentage of technology and business groups; whereas Boston showed to have a lower than average percentage of business groups. NYC had a slightly higher than average amount or both. At the sub-category, or thematic level, Boston showed to have a propensity of programming language groups, and a lower number of business networking groups.

Among the three cities, the most in-depth analysis was performed on Boston, and so these initial results will focus on this region. With respect to different recommendation components the preliminary findings included:

- Community-Community: The largest Boston groups have been software development and programming language. However, social network analysis revealed that a NewTech[8] group, comparatively a smaller and younger group, was a key group based on its centrality in the network. Geographic analysis was helpful to study why many small entrepreneurial groups were less connected to the core technology center (Boston has a well-known hub-and-spoke layout). The addition of zip code based data confirmed that suburban groups were less connected.

- City-City: Although a final goal was to have graph matching performed across entire regions, initial analysis was limited to sub-graphs, i.e. micro-clusters of communities. Supervised learning techniques were used find the nearest counterparts of groups in Boston versus those in Silicon Valley and NYC. For example, Silicon Valley did not have a "predictive analytics" group by name; though several interesting groups were discovered including "Graph Database Group" and "Big Data Analytics: Mobile, Social and Web". The "SV Business Intelligence" was the closest match.

- Member-City: Members included in the exploratory sample were Boston tech and business members who in their Meetup profile had provided a LinkedIN address AND listed their interests. In aggregate, the most frequently desired interests displayed were similar for both Meetup and LinkedIN. That is, member identities were consistent between platforms. The most desired themes included networking, startups, web, and social media (Meetup groups exist for each). Two of the highest unmet needs for Boston were for Perl and XML which for each the magnitude of interest was moderately strong. Next research steps will include sampling a broader selection of LinkedIN members.

---

[6] http://toreopsahl.com/

[7] http://gephi.org/

[8] http://www.meetup.com/Boston_New_Technology/

CONCLUSION

New data on community groups from Meetup.com has enabled more granular analysis for understanding how to both efficiently and effectively manage communities. Open questions remain as to the impact of these informal communities on local economic performance. However this phase of research was aimed at quantifying, and identifying new communities that may serve as initial recommendations, or inputs, to discussions on community development.

ACKNOWLEDGMENTS

REFERENCES

1. Saxenian, A. (1994). Regional Advantage: Culture and Competition in Silicon Valley and Route 128 (Cambridge, MA: Harvard University Press)

2. Powell, W., K.W. Koput, & L. Smith-Doerr (1996), Interorganizational collaboration and the locus of innovation: networks of learning in biotechnology, Administrative Science Quarterly 42 (1): 116-145

3. Putnam, Robert D. (2000). Bowling Alone: The Collapse and Revival of American Community. New York: Simon & Schuster.

4. Weinberg, B. D., Williams, C. B. (2006). The 2004 US Presidential campaign: Impact of hybrid offline & online 'meetup' communities. Journal of Direct, Data and Digital Marketing Practice, 8 (1 (July)), 46-57.