



Email Analytics for Customer Experience

A recent report estimated global email users to be 2.6 billion at the end of 2015. By 2019 the user base is forecasted to grow to 2.9 billion.

Introduction

Email is one of the primary modes of communication within an organization and between organizations. A report by Radicati Group calls email the “go-to form of communication in the business world”. A recent report by the same group estimated global email users to be 2.6 billion at the end of 2015. By 2019 the user base is forecasted to grow to 2.9 billion. The volume of business emails generated in 2015 was 112.5 billion emails per day and it accounts for little more than half of total worldwide email traffic. If we look at email as a data source, the volume and richness of data captured every day is enormous. Emails are archived for records, however, little emphasis exists on analyzing email contents and deriving useful business insights. This whitepaper aims to explore email as a data source and highlight emerging analytics techniques that can be applied to mine valuable information from email databases.

Email contains rich unstructured information in the form of content that users send to each other as well as meta-data associated with email system like sender/receiver information and dates which can be analyzed. Particularly, we look at two key aspects in which email communications are used – communication within an organization and external communication with customers. Internal emails can be analyzed to see communication patterns prevalent within organization and also monitor workflows which primarily rely on emails.

Patterns can also be analyzed at inter-team and inter-department level to give a high level overview. The insights thus gained can be used to improve operational efficiency. External communications can be monitored from the perspective of how an organization interacts with its customer and how customers perceive the products and services it offers. The paper begins with a brief introduction on analytic techniques which form the basis of more advanced use cases on each of two aspects described above.

Email Analytics using Social Network Analysis and Text Mining

Social Network Analysis is a broad term covering techniques to analyze *social structure*. A social structure is a network with *members* and *relationships* connecting the members. For example users who are friends on social media can be connected together to form a network. In the email context, an email network can be created by connecting the sender and receiver from the email meta data. *Figure 1* shows an example of network created for Hillary Clinton’s emails. The data was made public by the State Department on 13 Aug 2015. After creating a large-scale graph of email users in this way, algorithms from network theory can be used to find centrality or importance scores for each user which can then be applied to identify crucial email users like top contributors, influencers, bottlenecks etc. Email also contains rich unstructured data in its body.

Emerging techniques from text mining like sentiment analysis, topics extraction, and word co-occurrence analysis can be used on the content in emails, which can be used to gauge customer sentiment and key problem areas for customer.

In the subsequent sections, we describe the application of these methods on two sample areas – analytics for operational efficiency and for customer experience.

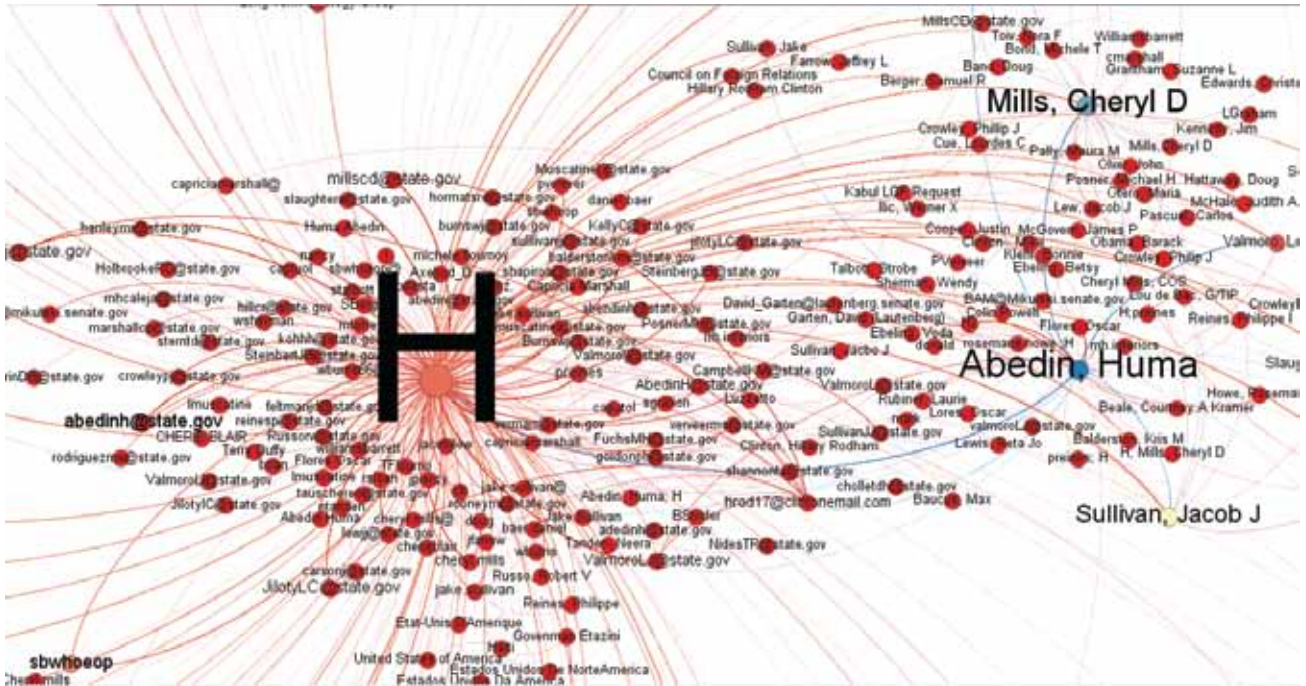


Fig 1: Network visualization of Hillary Clinton's emails. The central large 'H' is Hillary Clinton. The graph also shows Cheryl D Mills (Counselor and Chief of Staff) and Huma Abedin (Deputy Chief of Staff) as influential members in Hillary's network.

Analyzing email for Operational Efficiency

Operational efficiency is conventionally monitored through dashboards which track SLA adherence and performance indicators like customer satisfaction scores, first time resolution, time for resolution and others. Email data can augment conventional operational

dashboard by adding insights about team interaction aspects like engagement and collaboration which are otherwise difficult to quantify. In this section we discuss selected use cases to quantify such aspects.

Voluntary Leaders

Social network scores can be an important indicator of leaders within a team, both appointed and voluntary. Comparing this with actual organizational hierarchy, we get insights on voluntary leaders and people who assume leadership role without being officially assigned. A research¹ by authors at Columbia University was able to pick up most important individuals, group individuals of similar importance and draw a “social” hierarchy chart using Enron email dataset. Voluntary leaders is a closely associated concept with what is displayed in (Fig 1) in a sense that it brings to light who are secondary leaders in a network. “H” (Hillary Clinton) is the top influential member if we take the number of connections, but if we take other measures like *eigen vector centrality*, Cheryl D Mills and Huma Abedin come on the top.

Engagingness

Effective communications within a team is crucial for its success, and engagingness and responsiveness over email are key factors in making communication effective. A study² by researchers at Singapore Management University defines engaging users as those who effectively seek responses from others through their communication behavior, while responsive users are those who are willing to respond to communications sent across to them. The authors propose various models to quantify engaging and responsive behavior and derive corresponding measures. They use different models that use email threads, properties (like thread count) and email sequence gaps between their emails. Users’ social behavior is

modelled through social cognitive models which take into account the fact that people learn by observing people around them. This helps predicting which emails were responded to ahead of others based on engagingness and responsiveness of the users involved. Business can use the information to monitor engagingness and responsiveness of critical people in a workflow like team leader, single points of contact or customer facing users.

Collaboration

Collaboration is very important for quality output from teams. Since most of the data is shared and inter-team communication happens through email, it becomes an important source to quantify how effectively the teams are collaborating. We consider two key aspects to quantify collaboration from email data - one is contribution from all members involved and second is volunteering behavior shown by members. An important measure for member wise contribution is the *contribution index* which is calculated using number of emails sent and received by a user. A paper³ from MIT and Dartmouth College analyzed open innovation network (W3C) on the web and was able to segment the volunteers into web enthusiasts who work on open source projects as hobby, commercial contributors and academic contributors just based on data.

Community Detection

Network clustering techniques can be applied to the email networks to uncover underlying *communities* in the user base. The communities are group of users who send emails frequently to

¹ Creamer et al., “Segmentation and Automated Social Hierarchy Detection through Email Network Analysis.”

² On et al., *Engagingness and Responsiveness Behavior Models on the Enron Email Network and Its Application to Email Reply Order Prediction*.

³ Gloor et al., “Visualization of Communication Patterns in Collaborative Innovation Networks-Analysis of Some w3c Working Groups.”

each other, irrespective of whether they belong to the same team or not. A group⁴ affiliated with Technion, Google and IBM Research was able to accurately identify communities of project groups, guide-supervisors, academic course-staff and even a jokes forwarding group by mining emails. In a business context, this can be used to monitor compliance. By monitoring communities that have access to critical information, any unauthorized access can be brought to light. A good example is a study⁵ performed on Enron email dataset. This dataset was made public after the Federal Energy Regulatory Commission concluded its investigations in 2003. Findings from the study states that during time of the Enron crisis, the volume of communication intensified, and the organization was divided into distinct sub groups. After about two months since the crisis came to light, the strength of sub groups reduced as previously isolated users also started communicating frequently. It also suggests some interesting findings like the board reduced all communication to its lowers ranks while lateral communication within ranks like executive management increased.

Analyzing emails for Customer Experience

Engagingness and responsiveness described in the previous section become even more critical when the emails are directed to customers. By tracking these metrics for customer facing users, we can ensure that customers get the desired experience over email. In addition, the next sections describe use cases of predicting client satisfaction and analyzing feedback in form of complaints.

Email Communication and CSAT

By monitoring email that goes out of an organization, it is possible to observe communication patterns with customers. By correlating actual customer satisfaction score with social network metrics, a study⁶ by an author at Zeppelin University, Germany was able to find most significant factors in predicting client satisfaction. The study posts some interesting findings like - overall centrality scores and contribution index in group correlate positively with customer satisfaction, while average new team members added and leadership rotation impacts it negatively. A communication score card is proposed which can be monitored for communications going to customers. A sample score card is shown below which uses social network metrics from our own customer 360 solution HyperGraf™.

Complaints Analysis

SNA Metric	Direction of correlation
Group Knowledge Broker Score	+
Avg User Reach	+
Community Formation	+
Closeness Centrality	-
Leadership Oscillation	-
Sentiment	-

With email being a prominent medium for customer support, the contents of emails can be used to mine information about recurring complaints. Text mining methods like topic extraction and word co-occurrence analysis fit this scenario well. A paper⁷ from Monash University describes how the researchers found concepts by clustering words from a free text description. Then they link the concepts that are

⁴ Bar-Yossef et al., "Cluster Ranking with an Application to Mining Mailbox Networks."

⁵ Diesner et. al, "Communication Networks from the Enron Email Corpus: "It's Always About the People. Enron is no Different."

⁶ Brunberg, Gloor, and Giacomelli, "Predicting Client Satisfaction through (E-Mail) Network Analysis: The Communication Score Card."

⁷ Chau et al., "A Conceptlink Graph for Text Structure Mining."

related using a *network* to form a concept link diagram. In our scenario, such concepts can be problem areas that customers mention in the emails and any associated product features and related key terms. Such a diagram can help in root cause identification of the problems and point towards action areas. We used a modified version of algorithm described in the above paper to analyze US banking related complaints data publicly available at CFPB website. (Fig 2) shows a concepts network derived using that data. It displays words like ‘account’, ‘access’, ‘agents’ etc. together as a concept, connected with other concepts like ‘site web’ (web site), ‘comp’ (comparables in real estate). An analysis like this can give insights into espoused processes, technology bottlenecks and other operational issues. For example, if we investigate the ‘site web’ and ‘access account...’ link further, we find an example where a customer was unable to upload documents on website, and the agent was further not able to help as his office closing hours were approaching. Another illustration is a use case⁸ where the associated consultants

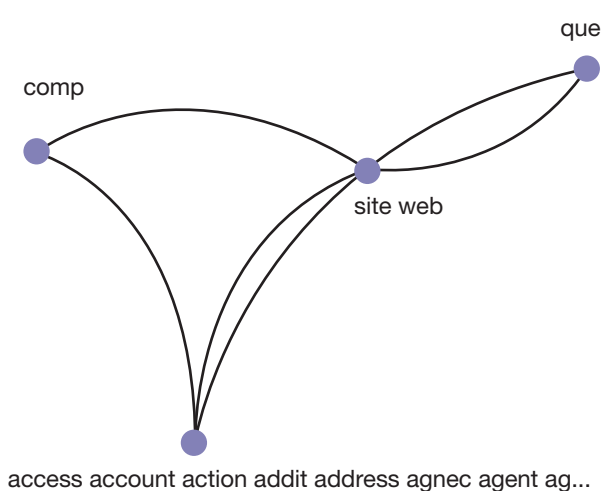


Fig 2: An example word network created for bank related complaints. The data was sourced from publicly available database at CFPB.

⁸ Chakraborty, Pagolu, and Garla, *Text Mining and Analysis*.

used emergency room data from hospitals pertaining to injuries and were able to identify key reasons that lead to injuries from falls and sports injuries.

In a nutshell, we described select use cases that showcase how email data can be leveraged to get insights on various *actors*' behavior. Using emails internal to organization we can quantify aspects like engagingness, collaboration and volunteering behavior in actors internal to organization like customer support teams. Email going to customers give insights on factors influencing CSAT and nature of customer complaints. Email data can thus be effectively used to enhance customer experience by taking appropriate actions on behavior of both these type of actors.

Conclusion

Organizational email is archived for records, but the data therein is not often analyzed for insights. Email being the primary mode of communication in organizations, insights from email data can uncover valuable information about communication patterns existing within and external to organization. In this whitepaper we have shown several possibilities of how this information can be used to monitor organizational health, improve operational efficiency and monitor customer interactions and feedback. Email analytics comes with two key challenges – the unstructured nature of data and volume of data generated. However, with tools like advanced algorithms for unstructured mining and big data platforms like Hadoop and Spark, these challenges can be effectively addressed.



Vibhav Kamath
Manager, Mphasis NEXTlabs

Vibhav currently works as Manager with Mphasis NEXTlabs. His areas of work include optimization, system dynamics, natural language processing, social network analysis, predictive modelling and cloud, and has applied it to areas like workforce planning and customer profiling.

Vibhav holds a Masters in Industrial Engineering and Operations Research from IIT Bombay and B.Tech in Electronics and Telecommunications from College of Engineering Pune. He has worked on 'discrete event simulation' in his Masters thesis project.

About Mphasis

Mphasis is a global technology services and solutions company specializing in the areas of Digital and Governance, Risk & Compliance. Our solution focus and superior human capital propels our partnership with large enterprise customers in their Digital Transformation journeys and with global financial institutions in the conception and execution of their Governance, Risk and Compliance Strategies. We focus on next generation technologies for differentiated solutions delivering optimized operations for clients.