Fragmentation of identity through structural holes in email contacts

danah boyd, Jeff Potter, Fernanda Viegas

This is the first half of a paper submitted for Keith Hampton's "Social Networks" course; the second half is focused on the visualizations and will be added later, due to obfuscation issues.

Abstract

Burt (1993) suggests that there are a number of advantages to maintaining structural holes in one's social networks, including controlling access to resources and maintaining personal privacy. Since structural holes segment an individual's social network into unconnected clusters, the individual is able to portray a socially appropriate facet of hirself[1] to each cluster separately, without feeling constrained by the combined social norms. This is significantly advantageous for marginalized individuals who fear retribution should certain aspects of their identity be made available in other contexts. In order to continue to maintain separate social personas, an individual must also explicitly maintain the developed structural holes.

While structural holes can be simply maintained in the physical world by associating particular physical environments with particular ties, this does not directly translate to the digital world. Due to aggregation of persistent data across most digital environments connected via search engines, location becomes meaningless in the virtual world because it's impossible to keep digital contexts from converging. Multiple online personas, usually maintained via separate email addresses, present a temporary solution. By associating a particular facet of one's identity with an email address, and maintaining all appropriate social ties via that address, an individual is able to build a context around an email address. Not only is this time consuming, but it's also a hassle since most systems are built to help you eliminate multiple contexts. Yet, even with the inconvenience, the opportunity for control of personal information motivates many to explicitly manage multiple addresses. As

---

[1] When reflecting on identity, gendered pronouns convey a tremendous amount of unintended meaning. At this juncture, there is no consensus on which non-gendered pronouns are most acceptable, although there are approximately 25 different publicly used variations. For the purpose of this paper, we have chosen to use a set that is the most comfortable and least derivative of traditional pronouns – s/he, hir, and hirself (all of which combine the traditional male and female pronouns).

Miller, Casey, and Kate Swift. 1980. *The Handbook of Nonsexist Writing For Writers, Editors and Speakers*. New York: Lippincott & Crowell.

Williams, John. 2001. "Gender-Neutral Pronoun FAQ." Version 0.9.10, October 29. http://www.aetherlumina.com/gnp/

a result, it's not uncommon for someone to copy/paste content in order to send the same message to different groups of people under different messages or with different email addresses. Doing so not only eliminates potential cluster convergence, but it also allows individuals to interact in a socially appropriate fashion.

In this paper, we will draw from theoretical ideas surrounding identity and social networks to explain how structural holes are advantageous for giving individuals' control over their identity. Stemming from this, we will briefly consider what implications the digital world has on the maintenance of structural holes, emphasizing how currently technologies encourage convergence. Referencing personal anecdotes, we will discuss how and why some people use email to maintain structural holes. To further articulate this behavior, we will present a series of visualizations that we built to observe one individual's social network as seen through his permanent email archive. By addressing the relationship between multi-faceted individual identity and structural holes from a variety of vantage points, we hope to build a basic foundation for more substantial work in this area. This paper presents initial strides in pursuing these ideas.

Introduction

The roles that we play and the social networks that develop around them help us define our individual identity. By reacting to the events and people in our lives, we see ourselves in relation to others and we use our reactions to shape our own sense of self. In contemporary society, many people play a wide variety of social roles – i.e. employee, mother, friend, sister, lover, and organization member. Based on the context of these roles, we interact with a diverse set people in a wide variety of different physical environments. Because of this diversity, our social networks are rarely fully integrated, either intentionally or not. For example, in many people's lives, work colleagues do not know family members, even if they are aware of one another.

As Burt (1993) articulated in his structural holes theory, the fragmentation of social circles offers us a variety of social advantages, particularly by giving an individual negotiation powers. The theories developed in social networks, particularly those surrounding the argument of structural holes, offer a valuable perspective for reflecting on individual maintenance of a faceted identity.

This paper intends to initiate such a reflection. First, we will introduce the prominent issues in thinking about a multi-faceted individual identity. Coming from an individual-centric perspective, we will offer a simple overview of related social networks research, emphasizing theories related to structural holes. Building from this, we will offer our own perspective on how structural holes give individuals freedom to maintain their identity in a desirable way.

Using the digital realm as our playground, we will explore how such behaviors can be observed in the social networks that individuals maintain. Coming from a position that

digital social networks are not inherently different than physical ones (Wellman, et. al., 1996), we will intermingle the two realms, noting specific cases when one impacts the others. We will use a series of observations about individual email habits to illustrate our ideas. In addition, in order to ground the conversation in actual behavior, we will analyze one individual's digital social network. Using his email archive, we developed a series of visualizations intended to explore his social network and other digital behaviors. In addition to analyzing the visualizations, we are using the visuals to initiate a more ethnographic conversation about this individual's behavior.

Although approached as a completed paper, the research presented here is not final. We have not finished developing the tools necessary to do a complete analysis and we've only begun the mini-ethnographic conversation about the data. As such, we will do our best to present these ideas, explain the weaknesses and where we are headed with this work. In a sense, this paper serves as a cross between a final paper and a proposal. Regardless of its intention, we feel as though this document serves to initiate the desired reflection surrounding the control of individual identity via maintenance of one's social network and specifically one's structural holes.

Theoretical and Conceptual Background

Although personally constructed, one's identity is impacted by one's interaction with others. Many people have a variety of roles in an individual's life and therefore they provide a variety of impact, ranging from the stranger on the bus to one's best friend. Not only does the strength of an individual's connection to others play a role, but also the context, the value and a wide variety of uncontrolled events. These people help comprise an individual's social network, or the collection of people of whom the individual relies on for a variety of purposes. Although these people surround the individual, they may not all know or even be aware of one another. These holes in awareness or knowledge can be described as structural holes within an individual's network, where the only relationship that one person has to another is through the ego whose network is being considered.

In order to consider the impact of such structural holes on an individual's sense of self, we must begin by introducing prominent theories of identity and social networks. In our introduction to social networks, we will focus the previous work on issues directly related to identity, therefore also introducing the relationship between identity and networks, as well as issues from the physical and digital domains. It's important to note that a large quantity of social networks theory is concerned with maximizing an individual's access to information, money, and power. Although we are using this theory, we have extracted aspects that are particular to individual control, regardless of usage intention.

Another weakness of the social networks literature is that it is mostly concerned with overdeveloped nations, wealthy cultures and professional individuals. As we are placing our examples within a digital context, we recognize that much of what we are assuming implies

participation in the same society that most of the literature addresses. As such, we assume individuals have access to a diversity of people, live in a society where most people 'go to work' and have other articulated societal roles, and have the physical/digital mobility to interact with different groups of people. For the sake of conceptualization, the modern day metropolis exemplifies these expectations.

Understanding individual identity

In modern society, an individual's self-awareness allows hir to have a sense of self in relation to society and culture. Such a self can be referred to as individual identity, a constructed and evolving reflection of the self, reflexively developed by social interaction. One's individual identity is comprised of two facets where one is internalized and the other is projected. Although there are numerous variations on these components (i.e. Smith's object self/acting self, Mead's me/I, Freud's ego/id), we will refer to these attributes as one's internal identity and one's social identity (Freud 1974 [1923]; Mead 1934; Smith 1976 [1790]).

Internal identity, or self-identity, refers to the self-perception of the individual in relation to the world. As it is reflective in nature, self-perception cannot be purely manifested internally. Without society as a basis for reflexivity, there can be no internalized evaluation. As such, social experience and interaction provides the model by which individuals can give meaning to the physical, psychological, philosophical, and moral aspects of their identity.

Alternatively, when an individual interacts with others, s/he conveys aspects of hirself through a set of signals that others must learn to read and evaluate. The negotiation between self-presentation and external evaluation can be viewed as a performance, which helps construct an individual's social identity (Goffman 1956). While individual identity is entirely constructed by the individual, social identity is perceived externally, relying not on the intention, but the effective expression and perception of an individual's presentation.

Through constant interplay, the social identity feeds the internal identity, which in turn manifests itself as an evolving social identity. Because a variety of contexts affect individuals differently, an individual's social identity appears to constantly change according to the social situation. Postmodern theory suggests that an individual's contemporary identity is fragmented, thereby producing an identity crisis for the modern individual. Alternatively, we believe that an individual's identity is simply multi-faceted[2]. While one's social identity is fragmented and constructed according to context, hir internal identity is cohesive and completely controlled. By (un)consciously understanding a social scenario, individuals perform a facet of their identity that is appropriate to the context and for the impression intended; likewise, the social forces that one integrates into hir identity are also impacted by the perceived value of the social context in comparison to the personal information.

---

[2] The groundwork for this argument is partially detailed in another paper that danah has written on modernity: http://www.danah.org/classes/modernity/ModernityFinalPaperSubmission.pdf

Social forces do not impact all people equally. In his discussion on social motivations, Snyder (1974) suggests that individuals have different styles of reacting to social expectations, or self-monitoring. Those individuals who are highly attuned to the expectations of others are considered to be high self-monitors while those who focus on presenting themselves regardless of social feedback are perceived be low self-monitors. Because of their attention to social cues, high self-monitors are more likely to facet their identity to fit particular contexts.

When individuals are able to portray different facets of themselves in different contexts, not only are they able to structure their privacy, but also they are able to develop socially appropriate personas, build context-dependent reputations, and otherwise engage with others on the expected level. For example, the identity information that one presents Saturday night at a bar is drastically different from the information presented at work. Not only do the different contexts expect different behaviors, but also the expected relationships and valued information are drastically different. Additionally, if participation at the Saturday night bar would be controversial at work, keeping this information private becomes more important. Such privacy concerns are particularly important to marginalized or non-mainstream individuals (i.e., the queer community).

Individual identity information stems from the facets that individuals show, either intentionally or unintentionally. Identity information can range from demographic information that is often readily accessible (age, sex, address, etc.) to more personal characteristics that define an individual's sense of self – historical experiences, values, practices, philosophies, etc. Practically any information about an individual can be considered part of their identity information, yet some aspects are more valuable than others. The more valuable the information is to an individual, the more likely that s/he will want to restrict who gains access to it. While value is self-determined, information that is likely to have severe consequences in particular environments is often highly valuable. For example, if an individual fears that s/he will lose hir job if hir employer learns about hir sexuality, s/he becomes highly motivated to restrict access to that information. Because marginalized individuals have genuine reasons to believe that revealing any aspect of themselves that differs from the norm will further oppress them, they are highly motivated to minimize the perceived differences; this is called assimilation. When an individual's social identity is only a small fraction of hir personal identity, and the remaining qualities are not socially normative, s/he must go to great lengths to limit access to those aspects. Because assimilation is just a social appearance, these individuals must either keep a great deal to themselves or systematically control who can gain access to this information.

Although the relationship between the personal and the social identity is crucial, for the purpose of this paper, we will focus predominantly on one's social identity. It is the social identity that is constantly faceted to meet the needs of an individual in a particular context. It's important to note that context refers to a variety of factors, including the features of the space, the historical interactions in that environment, the intended use of the space, and

the people within the environment. Awareness of these attributes gives an individual an idea of what's appropriate, thereby influencing hir presentation of self, or social identity. Yet, none of the physical attributes of an environment are as important as the people within. When an individual is performing aspects of hirself, s/he is focused on the perception of others.

The structure of a social network

By interacting with people, an individual starts to develop a community of people with whom s/he associates. Known as hir social network, this collection of people provides an individual with various degrees of support, potentially including economical and emotional support. The types of connections – or ties – that an individual maintains varies, but they often include family, friends, colleagues, and lovers.

Additionally, each tie may vary in value, or strength. Most commonly, social network theorists refer to two levels of ties – strong ties and weak ties, where a strong tie is able to offer a much greater magnitude of support than a weak tie. Although it may seem as though weak ties are not particularly valuable, Granovetter (1973) shows that there are distinct advantages to having weak ties, including increased information flow and social mobility. Since weak ties require less to maintain, it is in an individual's best interest to maximize hir weak ties, if s/he should want increased access to information.

In most cases, an individual has great control over the structure of hir social network. Although there are times and places when societies are so small or so tightly integrated that everyone knew everyone else, for many people this is not currently the case. More likely, individuals will develop associations with people who aren't even aware of most of the people in an individual's collection of acquaintances. These ties have a variety of purposes, and with each purpose, they have a difference in strength or importance. In some cases, a new tie might be neatly integrated in one's previously formed cliques. In others, that tie will be kept completely separate or only introduced to a limited number of one's ties. By controlling who knows who, an individual is able to explicitly manage hir social network, providing connections as s/he sees fit. When clusters of one's network are kept separate, a series of holes in the network develop, such that the individual becomes the bridge between the clusters; this is known as structural holes.

Burt's structural holes argument (1993) builds on Granovetter's weak ties argument (1973). Burt argues that the advantages of weak ties are magnified for an anchor who is connected to different social clusters which have no other bridging connection. In other words, an individual who is the only person connecting one clique to another is advantaged. Not only does the individual gain from having access to a different set of information, s/he has the power to control what aspects of this information can be shared with the different social clusters to which s/he belongs. Burt's discussion of structural holes is heavily motivated by the flow of social capital and the competitive advantages of controlling information flow. In his scenario, maximizing and controlling the flow of information is

essential and empowering, such that an individual seeks to acquire as much information as possible. Therefore, by being the bridge between multiple social clusters, an individual maximizes hir ability to acquire and control information. Although Granovetter argues that all bridges must be weak ties (1973), Burt rejects the relevancy of tie strength, but emphasizes that weak ties in bridges are more advantageous.

While Burt suggests that being a bridge is purely advantageous, Krackhardt (1999) argues that it's also constraining for the individual who acts as the bridge. In his analysis of cliques, Krackhardt develops the idea of Simmelian ties, where an individual is Simmelian tied to another if they 1) have a strong tie to one another and 2) share at a strong tie to at least one other person in common (i.e., they are part of a clique). Individuals who are members of a clique are constrained by the social norms of that clique such that Simmelian triadic ties are more constraining than simple, dyadic ties. Since each clique has a series of social norms by which its members are expected to follow, Krackhardt concludes that an individual who is a member of two separate cliques is constrained by the social norms of both groups, thereby needing to find the intersection of those norms in developing a socially acceptable persona. Rather than seeing the bridging role as empowering, Krackhardt views it as a restrictive position, except in the case of private behaviors. In private scenarios, where only the particular clique and ego know about the behavior, the ego is advantaged by being the bridge, because s/he can act differently in different groups. Thus, if an individual seeks to maintain different social behaviors in different contexts, s/he becomes motivated to control social situations such that two cliques cannot converge, thereby guaranteeing private scenarios.

It is precisely these private scenarios that an individual desires when s/he wants to maintain a multi-faceted individual identity. The individual produces hir own identity information; therefore its initial flow comes from its creator and s/he controls its initial recipients. Although trust and motivation plays a significant role in the passage of personal data, connections are also important. Regardless of trust and motivation, if information is passed to an individual with minimal ties, it's unlikely that the information will spread far. For this reason, one's social network is a considered factor when valuable private information is being shared.

From an individual's perspective, personal information is exceptionally valuable and therefore the individual wants to control its spread and content. The more valuable the information, the more closely the individual wants control. Should valuable information spread, it becomes gossip. Although individuals who are far removed are less likely to care to continue to spread the gossip, they are also less motivated to suppress its spread, as trust is less likely to override one's desire to spread information. In order to account for the potential of gossip, it's in an individual's best interest to minimize the ways in which it can spread. The most obvious mechanism is to only share information with ties who are close, trusted, and have no motivation to share the information. Another effective approach is to minimize access by developing and maintaining structural holes. Structural holes provide

security by 1) limiting the percentage of your social network that can learn any bit of information from other members; 2) increasing the number of degrees necessary for information to jump from one clique of associations to another. Although Milgram (1967) shows that minimal degrees are necessary to connect any one individual to another, by assuming that non-ties are less likely to continue the spread of gossip, increasing the degrees of separation effectively limits the passage of information.

Since flow of identity information can be more easily managed in a network with holes, it's in an individual's interest to maintain structural holes whenever possible, particularly when different cliques have different social norms. By being the only bridge between a set of work colleagues and a set of friends, an individual can portray two distinct social identities. Yet, once this faceting is started, it becomes more crucial that the structural holes are maintained. From Burt's information flow perspective, an added bridge simply weakens the power of the original bridge. When segmented identity information is involved, an added bridge can be considerably destructive for the ego, depending on the potential impact of revealing unknown identity information. In both cases, an individual is empowered by being able to act as a sole bridge between two different social clusters, although for slightly different reasons.

Some individuals instinctively separate many of their social clusters, if for no other reason than to minimize restrictions and maximize privacy. Just as Krackhardt noticed, when social clusters are bridged and Simmelian ties are built, an individual's behavior becomes constrained because s/he must follow the social norms of both communities simultaneously. Likewise, when an individual interacts with two cliques simultaneously, hir behavior is effectively public, requiring a participation that will be appropriate for both forums. Although aggregated conformity might be expected for some individuals, Kildruff's (1992) earlier work implies that the impact of such convergence might be highly dependent on one's personal qualities, in particular hir self-monitoring style. Because high self-monitors are quite likely to be influenced by their social surroundings, it follows that these individuals will be constrained when presented with combined social cliques with different norms. Conversely, as low self-monitors are less likely to adjust to social expectations, converging social clusters might not be so problematic.

It's important to note that converging social circles not only increase potential information flow and restrict acceptable behavior, but they also automatically increase identity information knowledge by making each cluster more aware of the individual's network. Should a clique be associated with particular activities or interests, others are likely to assume participation or interest. Depending on the difference in values and interests between the two groups, this may not be problematic. For most people, homophily alleviates this concern, such that any clusters that an individual might have are likely to be very similar to the individual, and therefore likely to be similar to one another (McPherson, et. al., 2001). Problems are most likely to occur when an individual maintains a cluster of people whose similarities to hir do not overlap with the similarities s/he has to another

cluster. For example, converging one's "anti-corporate/Marxist/activist" friends with one's corporate colleagues not only constrains appropriate behavior, but makes each group aware of the individual's involvement in the other.

While most social networks literature is concerned with the physical world, Wellman, et. al. (1996) maintain that the same concepts are equally valid for those networks built and/or maintained in the digital world. While the theories remain the same, the ways in which people can manage and control their social networks are inherently affected by the strengths and weaknesses of the interaction paradigms possible when using digital tools.

Issues in maintaining digital structural holes

In the physical world, managing structural holes can be done with relative ease. At the highest level, an individual can prevent awareness by simply not mentioning other ties within hir network. This is particularly common when maintaining weak ties, since awareness of a person's entire social network is not regularly expected. Additionally, by simply controlling which people can share the same physical space, an individual can easily manage hir social network so that structural holes have minimal interaction with one another. The more explicitly an individual fragments hir network, the more disconcerting unexpected convergences are. Although coincidences do happen, most people alleviate this potential drama by associating particular locations with particular contexts, and therefore with particular people. Most uncomfortable convergences can be avoided by not cross-pollinating ties with unassociated contexts.

While location is a dominant factor in offline structural hole maintenance, the explicit nature of it does not translate to the digital realm. Online, it is relatively simple to switch contexts; one can easily be engaged in two different chatroom conversations with individuals from distinctly different social circles. While multiple windows give the impression of multiple contexts, the ease with which people can rapidly switch between multiple contexts results in numerous accidents. It is not uncommon for individuals to mistakenly send an email or instant message to the wrong person. Although mistakes are often harmless, we've spoken with numerous teenage girls who can recall horror stories of sending a message talking about 'Bob' to 'Bob' instead of the intended close friend. Because most of these stories involved gossip, love interests, or feuding friends, they resulted in humiliation or intensified fighting. Although the conversation is specifically private, most digital interfaces make it simple to unintentionally broadcast a message improperly, resulting in inappropriate public behavior with significant consequences.

In addition to minimizing the effort required to switch contexts, the persistent nature of digital interactions eliminates the assumed ephemeral nature of a conversation. In the physical world, a passing conversation leaves memory traces, easily overwritten by new encounters; in the digital world, archived conversations can be infinitely accessible, making it challenging to escape previous interactions.

Not only can archives be tools for rehashing past private conversations, they present immense issues for individuals who are trying to structure their social network. With archives, not only can converging cliques share stories, they can actually produce previous interactions for consideration. While mom presenting your friends with recordings of your first baby steps is only slightly embarrassing, imagine your ex-girlfriend offering chat log fights to your new love interest.

Suddenly, just as Paul was about to clinch the job interview, he received a visit from the Ghost of Usenet Postings Past.

Combined with archives, search engines magnify the possibilities for contextual convergence, exemplified by Google's 20-year archive of Usenet. In one digital anecdote, an individual who commonly posted to Perl newsgroups on company time and bondage newsgroups on personal time lost his programming job because his employer determined that his behavior made the company look bad. When archived data became searchable, anyone could simply input someone's name and see all of their historical Usenet posts (as well as any other archived web, listserv data). In this way, everyone's digital Saturday night public bar behavior became transparent on Monday to those at work. Although both the bar and all of Usenet are public domains, a simple search across all time and space is currently only possible in the digital realm. Just as most people don't assume that their employer is observing them at the bar, they didn't assume that their behavior might come back to haunt them.

Aggregation of individual information across newsgroups, bulletin boards, websites, and listservs provides for tremendous contextual convergence. Not only must an individual control for what s/he produces digitally, but s/he must also control for what others say about hir and what private conversations become public because one of the discussants decided to upload the archive to the web. This level of control is virtually impossible without completely removing oneself from the digital domain. Instead, in order to be socially acceptable, an individual must perform according to the social norms shared by all observing parties. Since this means that s/he must accommodate the entire world (a.k.a. universal assimilation), s/he becomes incapable of presenting any unique identifiers or faceting hir identity in any meaningful fashion. Needless to say, to do so is completely undesirable, if not absolutely impossible.

When individuals recognize the implications of aggregation, many of them choose to either become anonymous or explicitly construct multiple digital personas. Most frequently, people maintain multiple email addresses that cannot be linked to one another or to the associated person. Although this allows an individual to maintain facets of hirself, hir presentation is ultimately fragmented or contextually void.

Complete aggregation makes digital location meaningless for considering context. In a sense, email address or digital persona serves as a context substitute. Interactions and people become associated with particular personas. Structural holes are built and maintained by controlling which email address particular clusters can access.

Unfortunately, tools for digital interaction are better suited for converging data than maintaining its separation. For this reason, maintaining structural holes is both challenging and time consuming. To gain the advantages of multiple digital personas, an individual must not expose data that ties them together. Initially, this simply meant not associating multiple personas with unique demographic markers such as name and location. More recently, it's become clear that digital identifiers such as IP address and cookie data are being used to link multiple digital personas. Although this aggregate isn't automatically linked to an individual's name, cookie records associated with e-commerce behavior present that connection with little added effort. While this may appear avoidable, corporations and advertisers have significant motivations for being able to accurately coalesce segmented personas. Not only are profile management/convergent tools (i.e., Microsoft's Passport.NET, AOL's Screen Name Services, Sun's Liberty Alliance) becoming more difficult to avoid, current research project trying to identify unique human characteristics will inevitably become products for merging divergent digital personas. For example, imagine when computational linguistics research on idiolects[3] has matured; using the web as a corpus, such a system will be able to unify multiple personas simply by recognizing the common language patterns unique to one individual.

Structuring social networks via email

While issues surrounding public data significantly impact how one can control access to individual data, analyzing more private data – like email – highlights the ways in which individuals struggle to maintain structural holes in a meaningful way. Rather than appreciating the ease with which they can manage their social network in one context, individuals work against the system to create separate personas. Not only do they go through the hassle of maintaining separate email accounts, they systematically copy/paste messages to multiple recipients under multiple messages rather than allowing one message to connect fragmented parts of their social network.

Multiple recipient emails (via the Carbon Copy function) are one of the best ways to quickly share a bit of knowledge with a wide variety of people. Yet, because of the need to be contextually appropriate and explicitly aware of one's social network, a mass email represents complete network convergence. Although the most obvious reason for re-mailing the same content under different messages is to give an individual the impression that the message is personal, we've observed a variety of other reasons in which users will

[3] Linguistically, an idiolect is the unique language pattern of a particular individual, evolved from the individual's local dialect. Chomsky (1986) built the philosophical foundation for this idea, motivating computational linguists to try to (dis)prove him.

re-mail a message. The following scenarios articulate just a few of the situations in which we've seen this behavior.

Scenario. Kate receives a message that she wants to forward to all of her friends because it is particularly humorous. Because she doesn't want her current boyfriend to know that she's talking to her ex-boyfriend, Kate copy/pastes the message first to her ex, next to her boyfriend, then to her group of friends, and finally to her work colleagues (using her separate work address). In this way, she's separated her lovers, let her friends know that they all received the message, and kept her work colleagues separate and properly associated with her work email address.

Scenario. After receiving a forward that entertained her, Claudia wants to send it to her group of friends. Although the message is in English, Claudia communicates with half of her friends in Spanish (although they all know English). When she forwards the message, she separates her group of friends into two groups based on their primary language of communication so that she can add her personal comment at the top of the forward in the appropriate language.

Scenario. Rob just returned from his vacation and wants to send an update about the adventures to his various friends, colleagues and family. Although he sends identical messages to his colleagues and family (the G-rated version), he separately sends those messages to separate those clusters. Additionally, he sends half of his friends the G-rated version and adds a few X-rated comments to the message that he sends to rest of his friends. Although the X-rated version is far more entertaining, he is well aware that its content is not socially acceptable to all of his friends.

Scenario. Hannah is putting together a large event and she wants to make certain as many people come as possible. In her email to announce the event, she explicitly tells her friends to pass it on to others that might be interested, particularly since she's sure she forgot various friends. She can't put everyone on the CC list because the list would be absurdly long and certain people shouldn't know that others were invited because they wouldn't come. At the same time, she doesn't want to blind-carbon-copy everyone because they wouldn't know which of their friends hasn't been invited. Instead, she chunks her address book into appropriate clusters of close friends, letting each group know that the majority of their clique has been invited, thereby helping Hannah figure out who still needs invites.

Scenario. Andre is sending a message to his mother about a fight that they've been having for a while. He wants his brother to be aware of the conversation, but he doesn't want his mother to know that his brother knows. Therefore, Andre includes his brother on the blind-carbon-copy list, making him aware of the message without letting his mother know. Andre is quite comfortable doing this since he knows his brother won't respond and reveal that he was included on the message.

In each of these scenarios, the people send a similar message to a group of relations, sometimes a collection of groups. The recipients are usually able to see who else received the message, which has a variety of different consequences. Effectively, the sender is bringing everyone on the receiving end together in one common space, making all aware of each other's presence. In Andre's case, his brother is effectively behind a one-way mirror, letting him see his mother but not vice versa. By choosing to control the recipient list in each case, each individual is attempting to structure hir social network in a way that seems appropriate for each individual context, as well as for group awareness. In the cases in which the sender chooses to use a separate email address, s/he's additionally associating a particular personal context with the social context.

Visualizing structural holes in a digital social network

In order to more clearly understand the issues of digital social network maintenance, we decided to explore the social connections present in historical email interactions. As many others have shown, email offers great insight into an individual's social network (Garton, et. al, 1999; Wellman & Hampton, 1999; Rice, 1994; Sproull & Kiesler, 1991).

In order to adequately analyze how email is used for maintaining structural holes, we felt that a complete historical record of all social connections would be necessary. Unfortunately, most people do not consistently archive all of their email interactions. Out of personal curiosity, one of the authors' friends – "Mike" – has maintained a complete archive of all messages sent and received during the last five years. Needless to say, we were overjoyed when he offered to let us analyze his massive data.

Given the enormity of Mike's email dataset and the value of visualization techniques in understanding large-scale datasets, we decided to develop a series of interactive visualizations to explore Mike's data. For inspiration and direction, we turned to previous visualizations of social interaction (Whittaker, et. al., 2002; Smith & Fiore, 2001; Sack, 2000; Donath, et. al., 1999; Donath, 1995). We felt that by constructing a visualization tool to suit our particular approach, we would be able to focus on understanding how, when and why structural holes developed. In addition, this gave us enough flexibility to adjust the parameters used to determine tie strength, add additional information on top of particular visual cues, and harness the value of motion in comprehending inexact data.

Although comparing multiple social networks would be extremely valuable, we were not able to find anyone with a similar magnitude of complete data. By using one individual's data to discuss structural holes in email networks, we are not suggesting that his data exemplifies generic digital behavior. Instead, we are using such an individually centric approach to gain insight into the problem, to offer credence for studying this behavior in more depth and across a wider variety of individuals. Another important advantage of this approach is that we are able to discuss the data with the subject, giving us the ability to ascertain the validity of certain observations, make sense out of outliers and acquire a

reading on the data that is not possible from a purely external vantage point. Such a directed approach allows us to understand where our technique is faulty before trying to generalize. Given that we are trying to gain insight into this problem, we feel as though this first step can only be beneficial.

Additionally, like many social visualizations, our approach is more valuable to those familiar with the individual or community represented. Although fascinating as a stand-alone piece, in conjunction with an understanding of the person, such visualizations emphasize patterns not otherwise considered. For example, Mike didn't realize that some people knew each other until he saw the visuals. Doubting the connections, he went back to his data; sure enough, the two people in question had engaged in a handful of conversations without Mike really taking note. For Mike, the feedback loop is even more valuable, effectively allowing him to see his social history. Given such a quantitative approach, Mike is able to view social patterns that were otherwise obscured during typical email interactions. Additionally, from a researcher's perspective, Mike's reactions to the visualizations are invaluable. Just by knowing his contextual relationship to his ties, Mike can shed light on patterns that might have otherwise gone unnoticed. Thus, in order to give context to our methodology and the resulting visualizations, we will first introduce Mike.

The subject: Mike

Mike is a 24-year-old, gay-identified, white male. Born in northern California, Mike moved to the east coast to attend Yale University, where he studied art and computer science. Subsequently, Mike has moved to Boston to work for a small company that specializes in technological art installations. During his undergraduate years, Mike worked for a variety of technology companies in California, Texas and New York City. In addition to these jobs, Mike developed many websites for himself and a few for various research psychologists. Alongside his technical work, Mike pursued his interested in various art forms, including printmaking, bookmaking and photography.

Mike's interests include movies, bicycles, artistic expression (bookmaking, printmaking, photography), media & technology & the web, psychology and cooking. Mike is currently single and temporarily on leave from his job. Mike's closest friends are dispersed across large cities in the United States, including Boston, San Francisco, New York City and Chicago. Mike keeps in touch with his friends through email, telephone, instant messenger, and by occasional visits. Mike gathers his friends regularly, either to catch a movie or for a night of homemade crepes; these events are frequently organized using email and web invites. On these occasions, Mike is quite likely to bring together a cross-section of his social network, so long as they are geographically collocated.

Mike has run in a variety of social circles over the years, some explicitly marked by temporary jobs/summer internships. A majority of Mike's social circles can be labeled as:

- Family
- High school friends

- Undergraduate friends
- Gay men in Boston
- Gay men outside of Boston
- Boston work colleagues
- Texas work colleagues
- California work colleagues
- Gay men in New York City

While many of these clusters were developed independent of one another, dominant themes in Mike's life can be seen throughout, implying significant homophily in his network. For example, a significant portion of each cluster is involved with technology and many of Mike's undergraduate friends are gay. Mike's generic tie closely resembles himself: a white, gay, male, educated, economically stable, technophile with creative tendencies.

Very little about Mike's life and identity is completely hidden; Mike is even out about his gay identity to almost everyone in his network, including his work colleagues and family. Mike is known to bring together groups of his friends, thereby opening the possibility for new bridges to be formed. At the same time, by observing Mike, it's apparent that he portrays slightly different characteristics depending on the situation. Virtually none of the practical seriousness by which he addresses work colleagues is maintained when he's interacting with his friends. Also, Mike's flamboyance increases exponentially relative to the percentage of gay boys in the environment; and his touchy-feely habits are reserved for close friends and potential interests. Although Mike can be solemn or cranky, his aura is typically one of a variety of levels of gregariousness.

Since 1996, Mike has collected all of the email messages that he has received and sent. Originally, he did this because it was easier to save everything than to determine what was valuable and what wasn't. Over time, Mike developed his own email program to manage the quantity of messages that he received. Because of this, he was able to manage an infinite number of email addresses. As such, he started assigning an email address per function. For example, work related folks had one address while school related folks had another. More specifically, when he helped gay communities with web work, he gave them a new address. This articulation of email addresses allowed him to quickly determine how the individual contacting him knew him.

Although Mike had barely heard of social networks prior to us presenting them, he was intrigued by the concepts and was excited to offer his network-related observations. For example, he was well aware of the fact that clusters of people had seemingly formed around particular email addresses, depending on the context through which they first met. We captured Mike's data prior to his knowledge of social networks, so as to guarantee that awareness wouldn't change his behavior, thereby affecting the characteristics of the data. Mike was also happy to help us analyze the data, giving us meaning behind the patterns we observed visually.

Method for analyzing Mike's data

For our visualization, we are using Mike's email data as the indicator of his entire social network. In order to discuss Mike's ties, we have categorized them.

Knowledge ties. We assume that if *A* sends a message to *B* that *A* 'knows' *B*. (We do not assume that *B* knows *A*).

Awareness ties. We assume that if *B* receives a message from *A* that *B* is 'aware' of *A*. Likewise, if *B* and *C* both receive a message from *A*, we assume that *B* and *C* are 'aware' of each other.

Trusted ties. If *A* sends a message to *B* and blind carbon copies (BCC's) *D*, we assume that *A* 'knows' and 'trusts' *D*. We assume this because *D* has the ability to respond and reveal that *A* included people without *B*'s awareness.

Additionally, we assume that most senders do not distinguish between the To and CC fields so we treat them identically (referred to as the To field from this point forward). We also assume that if no one is in the To field and everyone is BCC'ed that privacy is assumed and that there are no trusted ties. As we only have the messages that Mike received, we only know the people that he BCCs and the people that BCC him.

Example. Imagine the following messages:

> From: Mike
> To: Drew, Taylor
> BCC: Morgan, Kerry

This produces a set of ties as follows:

> Mike knows Drew; Mike knows Taylor
> Mike knows & trusts Morgan; Mike knows & trusts Kerry
> Drew is aware of Mike; Drew is aware of Taylor
> Taylor is aware of Mike; Taylor is aware of Drew
> Morgan is aware of Mike; Morgan is aware of Drew; Morgan is aware of Taylor
> Kerry is aware of Mike; Kerry is aware of Drew; Kerry is aware of Taylor

We maintain bi-directional links in order to understand the strength of ties. Using this information, we constructed a matrix of ties, including the quantity and type.

Based on these ties, we built a piece of software that would represent each individual as a node/anchor and each connection as a spring. All springs have an ideal length while individual springs have a springiness/weight dependent on a combined weighting of the associated knowledge and awareness ties. All anchors also have a magnetic repulsion to other anchors. When the system starts, anchors affected by strong springs connected to other anchors aim to converge, therefore pulling on all other connections. At the same time, springs are trying to reach ideal length and anchors are trying to repulse any anchor that gets too close. Over time, the system stabilizes into an optimal graph, minimizing the constraints and maximizing the goals.

Within our system, all constants (including the weights, repulsion, lengths and proportions) are changeable. We have yet to finalize the constants, as we are still exploring what combination makes the data most legible.

Unlike most individuals, Mike has developed a scheme by which he can group particular individuals based on the way in which he made a connection to them. By associating individuals with context-appropriate email addresses, Mike has made it tremendously easy for us to understand which aspects of his identity are associated with which people. For visualization purposes, we use the email address with which an individual is most frequently associated to color the individual's anchor. This allows us to visually see the impact of address association on clusters.

Dataset: Mike's email archive

Comprised of sent and received messages ranging from 22 March 1997 to 20 November 2001, the dataset consists of 80,941 messages with 15,537 unique people. Prior to 2000, no listserv messages were maintained, but since 2000, all messages sent and received were kept, including most listserv messages. Mike sent 24,064 messages and received 61,323 messages (with a 4,446 message overlap of messages sent from Mike to Mike for a variety of reasons, including testing his own email software, sending personal reminders, and posting to listservs of which he is a member). Excluding individuals who only send Mike message through a listserv, Mike knows or is aware of 7,250 unique individuals.

Awareness ties are not computed for listserv messages, nor are they computed for messages with more than 50 recipients. In the remaining data, we have constructed 662,078 ties. If we were to include awareness ties of messages with more than 50 recipients, the number of ties would explode to 11.7 million. Listserv messages that are also sent directly to Mike are kept as personal messages.

On average, a message has 1.03 recipients, suggesting that most messages are one-one communications. 7,336 messages have more than two knowledge ties (i.e., recipients); 4,134 messages have more than three knowledge ties. An average message has 8.18 ties. Note that a message has ties = 2 * S (N=recipients) N; such that a message with 1 recipients has 2 ties, a message with 2 recipients has 6 ties, and a message with 10 recipients has 110 ties. As such, this average suggests that a handful of messages have a large number of recipients and therefore ties.

Based on sent-messages, Mike has trusted ties to 266 individuals. Although the data suggests that Mike is trusted by 10,452 individuals, this is misleading; many of the messages that appear to BCC Mike, while including other recipients, are either spam or listserv messages. Yet, of the 266 individuals that Mike has trusted, 23 of them appear to trust him. This suggests that Mike's trusted numbers are correct and often corroborated.

By looking at knowledge ties constructed from messages originating from Mike, we can see that he knows 2,618 unique people. Of those, he has sent more than one message to 1,409 of them; and only 405 people have received more than 10 messages from Mike. This

suggests that the majority of people that Mike knows, he knows fleetingly (or that the knowledge tie is poorly defined). Additionally, Mike is aware of 4,632 unique people. From the other direction, we see that 2,620 individuals have knowledge ties to Mike while 3,921 have awareness ties to Mike.

[Sections are missing here, waiting for obfuscated data.]

Bibliography

Burt, Ronald. 1993. The Social Structure of Competition. Pp. 65-103 in Explorations in Economic Sociology, edited by Richard Swedberg. New York: Sage.

Chomsky, Noam. 1986. *Knowledge of Language: Its Nature, Origins, and Use*. New York: Praeger Books.

Donath, Judith, Karrie Karahalios, and Fernanda Viegas. 1999. "Visualizing Conversations." *Proceedings of the Hawai'i International Conference on System Sciences-32*, Persistent Conversations Track, Maui, HI: IEEE Computer Society, January 5-8.

Donath, Judith. 1995. "Visual Who." *Proceedings of ACM Multimedia '95*. San Francisco, CA, November 5-9.

Freud, Sigmund. 1974 [1923]. "The Ego and the Id," vol. XIX in *The Standard Edition of the Complete Psychological Works of Sigmund Freud (SE)*, eds. James Strachey et al. London: Hogart Press and the Institute of Psycho-Analysis.

Garton, Laura, Caroline Haythornthwait, and Barry Wellman. 1999. "Studying Online Social Networks." Pp. 75-105 in *Doing Internet Research*, ed. Steve Jones. Thousand Oaks, CA: Sage.

Goffman, Erving. 1956. *The Presentation of Self in Everyday Life*. New York: Doubleday.

Granovetter, Mark. 1973. "The Strength of Weak Ties." *American Journal of Sociology* 78: 1360-1380.

Kilduff, Martin. 1992. "The friendship network as a decision-making resource: Dispositional moderators of social influences on organizational choice." *Journal of Personality and Social Psychology* 62: 168-80.

Krackhardt, David. 1999. "The Ties That Torture: Simmelian Tie Analysis in Organizations." *Research in the Sociology of Organizations* 16:183-210.

McPherson, Miller, Lynn Smith-Lovin, and James M. Cook. 2001. "Birds of a Feather: Homophily in Social Networks." *Annual Review of Sociology* 27: 415-444.

Mead, George Herbert. 1934. *Mind, Self, and Society*, ed. C.W. Morris. Chicago: University of Chicago.

Milgram, Stanley. 1967. "The Small-World Problem." *Psychology Today* 1:62-67.

Rice, Ronald. 1994. "Network analysis and computer-mediated communication systems." Pp. 167-203 in *Advances in social network analysis: Research in the social and behavioral sciences*, eds. Stanley Wasserman & Joseph Galaskiewicz. Newbury Park, CA: Sage.

Sack, Warren. 2000. “Discourse Diagrams: Interface Design for Very Large Scale Conversations.” *Proceedings of the Hawai’i International Conference on System Sciences-33*, Persistent Conversations Track, Maui, HI: IEEE Computer Society, January.

Smith, Adam. 1976 [1790]. *The Theory of Moral Sentiments*, ed. D.D. Raphael and A.L. Macfie. Oxford: Claredon Press.

Smith, Marc and Andrew Fiore. 2001. "Visualization Components for Persistent Conversations." Pp. 136-143 in *Proceedings of CHI'01*, Seattle, WA, March 31-April 5.

Snyder, Marc. 1974. "Self-Monitoring of Expressive Behavior." *Journal of Personality and Social Psychology* 30:526-537.

Sproull, Lee and Sara Kiesler. 1991. *Connections: New ways of working in the networked organization*. Boston, MA: MIT Press.

Wellman, Barry, Janet Salaff, Dimintrina Dimitrova, Laura Garton, Milena Gulia, and Caroline Haythornthwaite. 1996. “Computer Networks as Social Networks: Virtual Community, Computer Supported Cooperative Work and Telework.” *Annual Review of Sociology* 22: 213-38, February.

Wellman, Barry and Keith Hampton. 1999. “Living Networked in a Wired World.” *Contemporary Sociology* 28(6), 648-654.

Whittaker, Steve, Quentin Jones, and Loren Terveen. 2002. “Managing Long Term Conversations: Conversation and Contact Management.” *Proceedings of the Hawai'i International Conference on System Sciences-35*, Persistent Conversations Track, Big Island, HI: IEE Computer Society, January 7-10.