

Implicit Factors in Networked Information Feeds

Fred Stutzman
School of Information and Library Science
University of North Carolina-Chapel Hill
213 Manning Hall
Chapel Hill, NC 27599-3456
fred.stutzman@unc.edu

ABSTRACT

In recent years, the development of "News Feed" interfaces have transformed the ways in which individuals seek and encounter information in social network sites. Rather than primarily searching for information, a networked information feed provides a constantly updated stream of information about ongoing activity in the networked community. In the following paper, the components of networked information feeds are examined. Particular attention is paid to the variable forms of content included in networked information feeds, the effects of diversity in network composition on the networked information feed, and the implications of filtering networked information feeds.

Categories and Subject Descriptors

H.3.3 [Information Search and Retrieval]: Information filtering, Selection process; H.5.2 [User Interfaces]; H.5.3 [Group and Organization Interfaces]: Synchronous interaction

General Terms

Algorithms, Measurement

Keywords

News feeds, information encountering, probability

1. INTRODUCTION

The growth in popularity of networked social interaction, through venues such as social network sites, afford new opportunities for peer-to-peer information sharing. Social network sites provide a set of novel tools that facilitate the disclosure of personal information to a group of alters, or "Friends" [2, 5]. Furthermore, social network sites employ intelligent information-generation techniques that can identify site activity to be shared as actionable information [4, 10]. The end product of social network site use can be thought

of as an *activity stream*, a collection of action in a networked system to be shared with alters as part of an ongoing information process.

According to boyd and Ellison, social network sites have three fundamental properties. First, they allow individuals to "construct a public or semi-public profile within a bounded system." Second, the individual is able to articulate connections to alters in the system. Third, these lists of connections can be "viewed and traversed" by others within the system [2]. Social network sites have traditionally been profile-centric, with information sharing generally occurring within the confines of a profile. In these systems, status updates, wall postings, shared links and pictures and other fundamental activities are centralized on the profile [2].

Ego networks in social network sites are characteristically large, particularly when compared to offline discussion networks [5, 8]. In early iterations of social network sites, "keeping up" with network alters required visiting multiple profiles, an inefficient process. To facilitate awareness of activity streams in the network, Facebook developed the News Feed, which "shows you all the actions your friends are making in real-time¹." Drawing on design principles of the email inbox and RSS reader, the News Feed is a centralized, real-time networked information feed of the activity streams of connected alters. The Facebook News Feed has proven highly influential, with similar feeds appearing at competing sites, including LinkedIn and Flickr. Furthermore, microblogging networks such as Twitter employ networked information feeds as principle interaction elements.

Situated at the center of a social network, the networked information feed has emerged as an important vehicle for information search and encountering. As networked information feeds are adopted at a wider range of sites, it is important for systems developers, designers, and researchers to understand the set of interactions that produce an individual's networked information feed. The purpose of this paper is to identify and explore the factors that contribute to the production, and differential experience of, a networked information feed.

2. CHARACTERISTICS OF NETWORKED INFORMATION FEEDS

An individual experiences a networked information feed (hereafter, "NIF") by viewing an stream of activity generated in a social network site (See Figure 1² for an example). The

¹<http://www.facebook.com/help/?page=408>

²The image used was provided as an exemplar by Facebook.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

HCIR '10 New Brunswick, NJ USA

Copyright held by author. ACM X-XXXXX-XX-X/XX/XX ...\$10.00.



Figure 1: Facebook’s News Feed is the canonical example of a networked information feed. ”Top News” represents a filtered NIF, whereas ”Most Recent” provides an unfiltered NIF.

NIF is a reflection of individual action in the system. For any time period t , let us define the contents of a NIF as the finite sum of content items c shared by alters, where an individual has alters a ($a = 0 \dots n$). An individual’s NIF for period t can be represented as:

$$NIF(t) = \sum_{a=0}^n c_a \quad (1)$$

where total (t) shared content is a function of network size and alter sharing behavior. An individual’s news feed for all periods T ($t = 0 \dots n$) can be represented as:

$$NIF(T) = \sum_{t=0}^n \sum_{a=0}^n c_a \quad (2)$$

where T is a fixed interval, and a and c are random variables. Therefore, the individual’s experience with a NIF is a stochastic process dependent on network size and the content sharing behavior of alters within the system.

In a site such as Twitter, the individual experiences the complete NIF, meaning content is not screened or filtered through any process. In sites where networks are large or multiple opportunities for content creation exist, individuals may be overwhelmed by information in a full NIF. Therefore, sites such as Facebook offer algorithmic filtering of the NIF. We can define the experience of the filtered (f) NIF as the finite sum of content items c shared by alters, where an individual has alters a ($a = 0 \dots n$), and content display is governed by a Binomial variable b . The Binomial variable b has two possible values $0, 1$ and a probability density function $p(f)$ specified by an algorithmic process. The functional form of $NIF(f)$ is therefore:

$$NIF(f) = \sum_{a=0}^n c_a b \quad (3)$$

In cases of filtered and unfiltered NIF’s, the real total of NIF activity can always be described by $NIF(t)$ and $NIF(T)$. When the NIF is filtered, however, the displayed total of NIF activity is described by $NIF(f)$. In the next sections, the first principles of content items c and network composition a are to be explored. The remainder of the paper focuses on factors of relevance in $p(f)$, the probability function governing the filtering process.

3. NIF PRINCIPLE ELEMENTS

The experience of a networked information feed is dependent on the size of the individual’s network of alters, and the content shared by the alters.

3.1 NIF content

In any NIF, there is a diverse range of content that can be shared: status updates, pictures, links, and much more [6, 10]. Within an NIF, there are essentially two types of content shared: that which is *individual-agentic* (c_i), and that which is *system-generated* (c_s). Content which is individual-agentic is intentionally shared by an individual through purposeful action; an example would be the posting of a status update. System-generated content is generally a report of a user’s action that is triggered by a system. Examples of system-generated content include NIF notices of ”friending” behavior, presentation of third-party conversations, or reports of event attendance (e.g. Figure 1).

Based on the simple ontology I have specified, an alter’s NIF content production can be described with the following form:

$$c(\text{alter}) = c_i + c_s \quad (4)$$

where c_i is a random variable representing individual-agentic content sharing, and c_s is a random variable representing system-generated content sharing for a finite in-

terval. We expect that both of these variables, c_i and c_s are contingent on u , a random variable describing site use by the alter. Therefore, the function describing total (T) expected alter NIF content production for all periods T ($t = 0 \dots n$) can be represented as:

$$c(T) = \sum_{t=0}^n \sum_{a=0}^n p(c_{it}|u_{at}) + p(c_{st}|u_{at}) \quad (5)$$

That is, the content production in a NIF depends on an alter’s actions in a site, which is conditional on site use. It must be noted that while c_i is completely dependent on an alter’s actions, it is possible that c_s will contain actions from the alter and the alter’s group of contacts. For example, if an alter is tagged in a picture, the system generates activity on behalf of the alter triggered by the actions of a third party. For simplicity’s sake, we do not create a third category, instead retaining third-party activity within c_s .

Previous work has identified Pareto, Zipf and exponential distributions for content production in large-scale socio-technical environments [1, 3]. Within the context of individual production, the probability of content creation is better specified with the Poisson distribution, defined as follows:

$$P(c) = \frac{e^{-\lambda} \lambda^c}{c!} \quad (6)$$

where c must be a non-negative integer³. Using simulation, we can explore potential impact of adding new individuals to the NIF at theoretical lambdas 2,4,6, and 8 (Figure 2). It should be noted that individual content production in NIF may also be well-specified with the Negative Binomial form (in the case of overdispersion). Furthermore, occasional or bursty network use may serve to zero-inflate alter NIF content production over time.

3.2 NIF network properties

The amount of content shared in a NIF is functionally dependent on the size of the alter network. Because we are able to specify an individual probability $p(a)$ of content production c by alters a (6), we are able to reasonably estimate the expected impact to an NIF of the addition of alters to the network with the following form:

$$E(c) = \sum_{a=0}^n p_a c_a \quad (7)$$

In practice, we find that the composition of an NIF network has a dual form. Similar to the ontology of content production, NIF network membership can either be *explicit* or *system facilitated*. Explicit network membership reflects intentional addition of an alter to a NIF network by an individual. Adding a Facebook friend, or “following” a Twitter user are examples of explicit network additions. System facilitated inclusion in an NIF refer to the penumbra of system-facilitated activities that allow non-members to participate in an individual’s NIF. For example, by “Retweeting⁴,” the alter employs system functionality to bring a potentially external individual into a NIF. Similarly, a comment left by

³It is my belief that the exponential features of power law-type distributions are incompatible with individual behavior, and therefore poorly specify these probabilities.

⁴i.e., hitting the Retweet button.

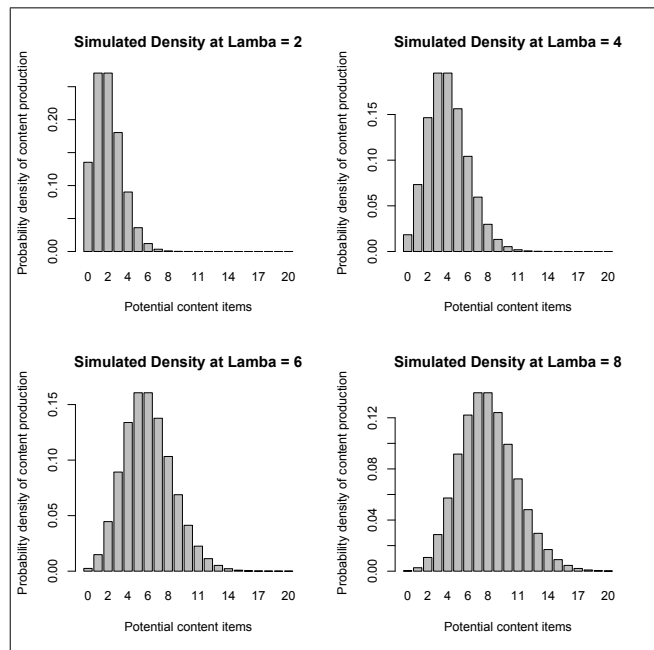


Figure 2: Simulated probability densities for NIF at theoretical lambdas 2, 4, 6, and 8

an alter on a potentially external individual’s status update in Facebook may bring the non-member’s content into the NIF.

The impact of adding members to an NIF network is more experientially complex than a simple addition of volume. We assume that the individual has relationships with content-producing alters, and therefore evaluates NIF content in light of these relationships. Furthermore, the particular configuration of relationships within an NIF will affect the individual’s experience of content. Drawing on theories of tie strength, and empirical models of network configuration, I briefly explore the experiential impact of network configuration.

3.2.1 Tie strength

An abundance of literature is devoted to the properties of social ties [7, 9]. The general principle is that each relationship has a certain strength, which is placed on a continuum of weak to strong. Represented mathematically, we can assume that each relationship an individual has with an alter can be represented by the random variable s , with possible values $0 \rightarrow 1$.

Although we are able to place a theoretical value on each relationship within a NIF, it is unwise to assume that “tie strength” always corresponds to a relevance judgment. Early work on tie strength identified the unexpected value of weak ties [7]; in an NIF, a person may allow weak ties into the network to observe ongoing activity, with only a small portion of activity having high relevance. Furthermore, algorithmic identification of relationship strength has proven to be challenging, though there is good work identifying metrics that may be more useful than others for identifying strength [6].

3.2.2 Network configuration

Within the bounded network of a social network site, we

are able to identify certain characteristics in relationship patterns that may influence the NIF experience. In a network, each vertex (or node) is connected to alters via an edge. The configuration of edges in the network are detectable through graph theoretic techniques, and may be useful for understanding individual experience of the network. Consider a simple measure of graph structure, density:

$$Density(\Delta) = \frac{2E}{v(v-1)} \quad (8)$$

In which the number of incident edges (E) is a proportion of the total vertices (v) possible in the graph [11]. A highly dense graph would indicate a network with many shared connections. A nuclear family, for example, might expect to have a highly dense network when replicated into a NIF. On the other hand, a sparse network indicates a heterogeneous mix of connections. A traveling salesperson, for example, may have many connections that do not share network edges. The density of the network, as well as the relative degree density between prominent members of the NIF has implications for the variety, relevance, and actionable nature of the content shared within the network.

Another important factor is the relationship between network configuration and content sharing behavior. When we specify the probability of content production (6), we assumed that the lambdas would be (approximately) normally distributed across the population (a safe estimate at large population sizes). At the subgraph level, however, it is possible that lambdas covary with network density, i.e. $cov(\lambda, \Delta) \neq 0$. This is potentially the result of local processes where some networks are incited share more than others as a result of influential, vocal few. Therefore, we assume that certain network configurations have characteristic NIF behavior and should be modeled with group variance.

4. THE FILTERING PARAMETER

Finally, we consider the NIF filtering parameter (3), which governs the display of content in a filtered NIF. As previously discussed, the filtering parameter is a Binomial variable b that has two possible values $0,1$ and a probability density specified by an algorithmic process.

It is beyond the scope of this paper to specify potential filtering parameters (see [6] for an extensive list of possibilities). In general, we assume that the designers of NIF filters want to maximize *interest* in the stream. We can therefore define max_i as a locally maximized vector of filtering variables $[f_1, f_2, \dots, f_n]$.

In a filtered NIF, we assume that all content items T (2) shared in an individual's NIF have a probability between $0,1$ of inclusion in the filtered NIF. Therefore, we define the log odds of inclusion of a content item to be:

$$\log \left[\frac{p_i}{(1-p_i)} \right] = \alpha + \beta_t x_{i1} + \beta_n x_{i2} + \beta_{[max_i]} x_{i3} + \beta_k x_{ik} \quad (9)$$

In this simple form, the log odds of content inclusion are a function of individual content production t , network size n , the local maximization vector max_i and a random term k .

5. CONCLUSIONS

Networked information feeds such as the Facebook News Feed are increasingly becoming an important place to both seek and encounter information. Situated in the midst of a social context, the NIF has the potential to continuously deliver relevant information from a large network of connections. This new form of information retrieval poses challenges to designers and researchers. How can the utility and interest of content shared in a NIF be maximized? What variables have the greatest potential to affect experience with the NIF? What factors are most important when filtering NIF's? This paper is the beginning of a research project aimed at answering these questions, which are of critical interest to industry, academia, and users of networked information feeds.

6. REFERENCES

- [1] A. Barabasi, R. Albert, and H. Jeong. Scale-free characteristics of random networks: The topology of the world wide web. *Physica A*, 281:68–77, 2000.
- [2] D. Boyd and N. Ellison. Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1), 2007.
- [3] A. Broder, R. Kumar, M. F., P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener. Graph structure in the web. *Computer Networks*, 33(309), 2000.
- [4] M. Burke, C. Marlow, and T. Lento. Feed me: motivating newcomer contribution in social network sites. In *CHI '09: Proceedings of the 27th international conference on Human factors in computing systems*, pages 945–954, New York, NY, USA, 2009. ACM.
- [5] N. Ellison, C. Steinfield, and C. Lampe. The benefits of facebook "friends:" social capital and college students' use of online social network sites. *Journal of Computer Mediated Communications*, 12(4), 2007.
- [6] E. Gilbert and K. Karahalios. Predicting tie strength with social media. In *CHI '09: Proceedings of the 27th international conference on Human factors in computing systems*, pages 211–220, New York, NY, USA, 2009. ACM.
- [7] M. Granovetter. The strength of weak ties. *The American Journal of Sociology*, 78(6):1360–1380, 1973.
- [8] M. McPherson, L. Smith-Lovin, and M. Brashears. Social isolation in america: Changes in core discussion networks over two decades. *American Sociological Review*, 71(3):353–375, 2006.
- [9] R. Putnam. *Bowling Alone : The Collapse and Revival of American Community*. Simon & Schuster, New York, NY, 2001.
- [10] E. Sun, I. Rosenn, C. Marlow, and T. Lento. Gesundheit! modeling contagion through facebook news feed. In *Proceedings of the Third International Conference on Weblogs and Social Media*, San Jose, CA, May 2009. AAAI Press, AAAI Press.
- [11] S. Wasserman and K. Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, New York, 1994.