

Chapter 5

Qualitative Analysis of Commercial Social Network Profiles

Lester Melendez, Ouri Wolfson, Malek Adjouadi, and Naphtali Rische

5.1 Introduction

Social-networking sites have become an integral part of many users' daily internet routine. Commercial enterprises have been quick to recognize this and are subsequently creating profiles for many of their products and services. Commercial enterprises use social network profiles to target and interact with potential customers as well as to provide a gateway for users of the product or service to interact with each other. Many commercial enterprises use the statistics from their product or service's social network profile to tout the popularity and success of the product or service being showcased. They will use statistics such as number of friends, number of daily visits, number of interactions, and other similar measurements to quantify their claims. These statistics are often not a clear indication of the true popularity and success of the product. In this chapter the term product is used to refer to any tangible or intangible product, service, celebrity, personality, film, book, or other entity produced by a commercial enterprise.

Social network profiles are complex entities that require one to use heuristics in addition to traditional statistics in order to assess the success and popularity of the product showcased in the profile. For example, a product's profile may have one million friends; using only quantitative analysis, this would lead one to believe that the product showcased has achieved a great deal of success. Using heuristic techniques one can sift through subsets of the profile's friend list and take into account things such as, the number of friends that are not other commercial profiles which interact at least bi-weekly with the profile being analyzed. This will yield a new measurement that may be orders of magnitude lower than the initial one million friends claimed. Thus, showing the success of the product was not as

N. Rische (✉)
NSF Industry/University Cooperative Research Center, Florida International University,
University Park, ECS-243, Miami, FL 33199, USA
e-mail: rishe@cs.fiu.edu; ndr@acm.org

great as initially thought when applying only quantitative analysis. Heuristic based qualitative analysis of commercial social network profiles is vital to understanding the impact, popularity, success, and reach of a product or service.

As with many academic avenues, there are outliers and caveats that do not fit neatly into the methods proposed; such instances are out of the scope of this chapter. This chapter will discuss some general methods used in traditional quantitative analysis of commercial social network profiles and introduce heuristics that will facilitate more accurate qualitative analysis in most cases. These heuristics have been developed with the help of partners in the music and entertainment industry.

5.2 What Is a Commercial Social Network Profile?

The social network profiles that the average user will sign up for are standard personal profiles. These profiles are designed for an individual user to use as their own public billboard. Standard personal profiles contain the features an everyday private citizen is likely to require. In addition to a standard personal profile many external social-networking sites, such as MySpace, offer their users the option of specialized profiles [1]. Specialized profiles are created with additional functionality – which is geared to accommodate the needs of a specific domain. These profiles may have additional features not available in standard personal profiles, while still maintaining most of the features that come with standard personal profiles. Commercial enterprises see these additional features as an opportunity to showcase their product in a dynamic and vibrant manner. Often, commercial enterprises will sign up for an account with a social-networking site and create a personal or specialized profile for the sole purpose of furthering their product’s success. The resulting profile is considered a commercial social network profile. It is important to distinguish commercial social network profiles from fan or imposter profiles.

Fan profiles, similar to fansites, fanpages [2], and tribute pages, are personal or specialized social network profiles created and maintained by supporters of a product for the purposes of paying homage. At times, fan profiles may be inadvertently confused with commercial social network profiles due to the level of devotion of the individual or group maintaining the profile. The content may be so involved and of such high quality that it is difficult to determine whether the profile is commercial or not. Imposter profiles pose a similar problem such that they are personal or specialized social network profiles created with the explicit purpose of falsely identifying themselves as a commercial social network profile. Imposter profiles often times are nearly identical in design and content to the commercial social network profile they are impersonating. It is difficult to determine the difference between the profile types mentioned without qualitative analysis.

Heuristic 1 provides a straight forward rule for determining whether or not a profile is commercial.

Heuristic 1:

- A social network profile created and maintained by a commercial enterprise for the sole purpose of furthering their product's success is considered a commercial social network profile (CSNP).

5.2.1 *Reciprocal Identification*

Application of heuristic 1 requires a qualitative analysis of the profile in question. Introduced here is a generic approach that can be applied across different social-networking sites.

An efficient and effective method for the application of heuristic 1 is reciprocal identification. Reciprocal identification (RI) is the presence of a direct hyperlink from the official website of the product or commercial enterprise to the corresponding CSNP. RI establishes the CSNP and the commercial enterprise's recognition of the relationship between the two.

RI can be detected by analyzing the hyperlink structure of the product or commercial enterprise's official website. Many times, it is necessary to spider the hypertext of the product or commercial enterprise's website in order to establish RI. It can be inferred that the profile is likely a CSNP if RI is found to exist.

The presence of RI can be established using automated scripts or manual inspection. An example application of heuristic 1 using RI is depicted below in Fig. 5.1.

Profile in Question:

<http://www.myspace.com/falloutboy>

Official Website of Commercial Enterprise:

<http://www.falloutboyrock.com>

Source Excerpt from:

<http://www.falloutboyrock.com>

```
<a class="footer_link" id="foot_ms"
href="http://www.myspace.com/falloutboy"
target="_blank">&nbsp;</a>
```

Result: RI is present because a direct link to the CSNP in question exists on the commercial enterprise's official website thus; <http://www.myspace.com/falloutboy> is a CSNP.

Fig. 5.1 Application of heuristic 1 using RI

5.2.2 *Types of Commercial Social Network Profiles*

One can categorize most social network profiles as commercial or non-commercial using heuristic 1 at a coarse-grain level. A fine-grain analysis shows that there are various subtypes of CSNPs. This chapter introduces five major types of CSNPs: music, film, television, public figures, and events. Determining which subtype a CSNP falls into requires a simple logical determination as to what product is being highlighted; the design of CSNPs makes this easy to determine.

5.2.2.1 Music

The music industry has been keen on using social-networking as a tool to expand the reach of their products; namely, performing artists. Market demand has made it almost certain that all music CSNPs are specialized profiles. Most social networks offer special functionality to accommodate the demands of music industry CSNPs.

As with most CSNPs, music industry CSNPs can be identified easily by commonsense qualitative observation of the CSNP itself [3]. Additionally, they share common content traits that can assist an inexperienced individual in identifying a profile as a music industry CSNP.

Examples of common content traits for music industry CSNPs are:

- Embedded multimedia player containing a sampling of the artist's music
- Hyperlinks for purchasing the artist's music
- Listing of artists upcoming tour itinerary
- Banners advertising the artist's upcoming featured appearances in media
- Banners advertising the artist's sponsors
- Links to promotional materials (banners, icons, wallpapers, etc.)

The facility provided by music industry CSNPs to present users with such a wide array of content, allows them to serve as stand-alone entities capable of negating the need for an official website if the commercial enterprise behind it so desires; though this is rarely the case. Generally, music industry CSNPs have a lifespan slightly greater than that of the artist they were created to promote. Once the artist becomes inactive, the CSNP may fall into disrepair and eventually be discarded by the commercial enterprise that created it.

5.2.2.2 Films

The motion picture industry has also found a use for CSNPs; these too are usually specialized profiles. Motion picture studios may create a CSNP for one of their upcoming releases [4]. These CSNPs are geared more towards fan-to-fan interaction. Many times the CSNP's main attraction is a forum where fans of the movie can interact with each other pre- and post-premier. The content offered is usually limited to trailers of the movie, content recycled from the official website, and links to the

official website. Motion picture CSNPs are used more as tools to funnel potential customers to the official website and establish a social network presence than to serve as a stand-alone attraction. This kind of CSNP has a relatively short lifespan. They rarely remain useful long after the theatrical or DVD release of the film it was created for.

5.2.2.3 Television

Television CSNPs are CSNPs built around products produced by the network and cable television industries. Similar to film CSNPs, they are used primarily to steer fans and potential fans towards the official website. Largely community based, these CSNPs rely on continuous fan-to-fan interaction and fan-to-product interaction to produce a return on their promotional investment; more specifically, creating a cult following and producing higher ratings [5]. Unlike film CSNPs, the lifespan of these profiles is expected to run long after the debut of the product; all the way through to the product's finale.

5.2.2.4 Public Figures

Public figures from television, fashion, sports, politics, and other avenues are known to use standard personal profiles to increase their visibility and to serve as a public relations tool. It is common for personalities to have a CSNP on all of the top social-networking sites [6]. Public figure CSNPs come in all shapes and sizes. Some are the public figure's primary internet presence serving as a means for press releases, media requests, appearance announcements, and more; while other public figures use their CSNP as nothing but a place holder to prevent squatters from creating imposter sites.

5.2.2.5 Events

CSNPs promoting upcoming events are now commonplace [7]. The commercial enterprises behind events such as the Super Bowl, the Vans Warped Tour, and the American Cancer Society's Relay for Life have all left their footprint on the social-networking world. These profiles are the shortest lived of the five subtypes of CSNPs as they are of little to no use once the event is over. They can be either standard personal profiles or specialized profiles. These CNSPs rely almost entirely on participant interaction. The sites have a narrow demographic as the reach of the event is generally confined to a small geographic area. Annual events may, at times, reuse the previous year's CSNP; updating it as necessary.

5.3 Quantitative Analysis of CSNPs

Quantitative analysis is the preferred method by which commercial enterprises evaluate the success of their CSNPs. They use statistics such as the number potential customers connected to the CSNP, customer interactions with the profile, and profile visits as selling points when touting the overall success of the product associated with that particular CSNP. It is important to have a fundamental understanding of some of the more commonly used measurements in order to grasp the importance of using a qualitative approach when analyzing CSNPs.

5.3.1 *Connections*

Profiles within a social network are often referred to as nodes within the technical community. Each node in a social network has a set of nodes that it is connected to via a social relationship. The nature of this connection, or relationship, is dependent upon the kind of social network. Some examples of connections are friendship, membership in an organization, ancestral ties, and geographic location. One of the more widely used types of connections is friendship.

For the examples in this chapter, it is safe to assume that friendship is the type of connection being used. If two nodes are connected to each other via a friendship relationship, it is said that they are friends. The set of all of a profile's friends is colloquially known as a friend list. In the realm of quantitative analysis, the size of a CSNP's friend list is a prime indicator of its success. Quantitatively, the conclusion can be drawn that the more friends a profile has, the more expansive its reach, and thus, greater its success.

For clarity, in the following scenario, the profile initiating the connection request is known as the "friender" and the profile receiving the connection request is known as the "friendee"; the connection request itself is referred to as a "friend request" and exists on the same social network. There are two ways for a profile's friend list to grow. The first is for the CSNP to act as the friender and initiate a connection with a potential customer, the friendee, by sending a friend request. This request asks permission to add the friendee to the friender's friend list and as a result adds the friender to the friendee's friend list. The friendee, the potential customer, now has the ability to accept or decline the friend request. There are various methods by which commercial enterprises go about initiating friend requests. Many times these friend requesting methods will skew quantitative analysis which will be discussed in greater detail in a later section.

The second is a customer initiated friend request. The potential customer submits a friend request to the CSNP. The CSNP's administrator then decides whether or not to accept the request. Once the request is accepted, the CSNP and the potential customer are friends and are present in each other's friend list. Once the potential customer and the CSNP are on each other's friend list, they are able to interact.

5.3.2 Interactions

By definition, social interaction is an interaction between two or more members of a social group [8]. Analogously, nodes within a social network can interact as well. Generally, the only requirement is that they are friends. Social networks offer a variety of ways in which their nodes can interact with one another. Interactions between nodes are singular one way events; meaning that one node is the sender and one node is the recipient of the interaction. Interactions are also either related or unrelated. Related interactions are interactions that are conducted in response to another interaction. An example of related interactions could be a private message thread in a user's inbox or series of comments made on two profile's main pages. The following is a sample transcript of four related interactions between a band's CSNP and the personal profile of a fan. The type of interaction being illustrated is a wall comment, also known simply as a comment.

- (1) **Band CSNP:** Hi there, are you coming to our performance tonight?
- (2) **Fan's Profile:** OMG! Yes! It's going to be awesome!
- (3) **Band CSNP:** That's great to hear = Will you be bringing friends?
- (4) **Fan's Profile:** Of course, my best friend Jane and her sister are coming too.

Interaction (1) triggers, or directly causes, interactions (2), (3), and (4) to be performed and thus all four interactions are considered related. Conversely, the following example shows three unrelated interactions.

- (1) **Band CSNP:** We have a new song up! Come check it out!
- (2) **Fan's Profile:** Hey, my friend says she went to HS with your singer.
- (3) **Band CSNP:** Don't forget, we have new t-shirts for sale!

A brief semantic analysis of the contents of these comments shows no discernible connection. None of the comments show any signs of having triggered the other and thus are considered unrelated.

Commercial enterprises place a lot of stock in the number of overall interactions associated with their CSNP. Quantitatively, it is inferred that the more interactions a CSNP has, the more interest its product has generated and thus the more successful it is. Examples of interactions may be related to the CSNP's main profile page, blog entries, status updates, pictures, videos, bulletin boards, and could even just be generic e-props, such as "kudos" or "likes".

The most popular type of interaction between social network friends is known as the wall, or profile comment [9]. The profile comment is simply a plain text or HTML snippet that the sending friend can leave on the receiving friend's main profile page. This snippet is displayed on a virtual corkboard of sorts on the receiving friend's main profile page. There are variations of comments implemented throughout the social-networking world; such as comments specific to pictures and videos. Though the most prevalent, comments are only one of countless interaction types.

Social networks offer many more novel interactions allowing friends to rank each other's pictures, videos, and blogs; send e-gifts, survey results, event invitations, birthday reminders, and more.

5.3.3 *Hit Counters*

As with their official websites, commercial enterprises find it beneficial to add hit counters to their CSNPs. Hit counters allow commercial enterprises to assess the traffic on their CSNP. Hit counters need not be limited to tracking the number of visitors who view the CSNP's main profile page. There can be specialized counters attached to specific blogs, videos, pictures, bulletin board posts, and just about any other element of a profile. Many social networks will provide users with options for attaching built in counters to various elements on the profile.

In the case of music profiles, social-networking sites such as MySpace and Facebook provide commercial users with the ability to keep track of fine grain statistics such as how many visitors to their profile listen to the music on their profile's embedded music player. Quantitative analysis takes into account the number hits a profile receives and simply put, it says that the more hits a profile has the more successful it is.

5.3.4 *Updates*

The number of updates made on a CSNP is not often used in quantitative analysis; but, nonetheless it is important as it will be used to cover qualitative analysis later in the chapter. The number of updates during a period of time can show whether a product has been abandoned by its commercial enterprise or whether it is being backed by an active marketing campaign. Updates are elements such as blog entries, status changes, tweets, bulletins, and others; and can trigger interactions as well. The number of interactions triggered by each update can also be a telltale sign of a product's success.

5.4 **Qualitative Analysis of CSNPs**

Quantitative analysis leaves much to be desired in the way of accurately assessing the success of a CSNP. The number of friends, interactions, hits, and updates can easily be skewed by any number of factors; such as spam, profile hacks, and automated scripts. Intentionally or unintentionally, the actions of the commercial enterprise behind the CSNP can greatly affect the accuracy of quantitative analysis. Many times commercial enterprises look to artificially influence their CSNP's stats in order to back up their claims of success. Additionally, the popularity of social networks has created a spam epidemic. The number of businesses that were targets for spam, phishing and malware via social networking sites increased dramatically, with spam showing the sharpest rise from 33.4% in April 2009 to 57% in December 2009. This highlights a surge in exploitation of such sites by spammers [10]. This creates the need to sift through any statistics skewed by spam and internal actions in order to arrive at a valid conclusion regarding a CSNP's success.

In order to discern between valid and skewed statistics it is necessary to understand just what factors can skew a CSNP's stats. The following sections will introduce heuristics that can be applied to raw quantitative data to produce an accurate picture of a CSNP's success. The heuristics introduced can be applied together or separately and the more of these heuristics are applied concurrently, the more accurate the results. It is up to the individual performing the analysis to decide what heuristics should be applied to suit their needs.

5.4.1 Connections

The size of a profile's friend list has long been the preferred measure of a CSNP's popularity. A clear application of this principle can be found in the music industry. Record labels, management firms, and concert promoters rely heavily on this measurement alone to validate their claims that the performing artist they are backing is genuinely notable and successful. Their reasoning is that each friend on the CNPS's friend list equates to one devoted fan. This trend is not limited to the music industry; film, television, public figure, and event profiles suffer from the same reliance on friend list size and the assumption that one friend equals one devoted fan as a measure of popularity and success.

5.4.1.1 Valid Connections

Obtaining an accurate count of the number of friends a CSNP is connected to goes further than simply looking at the size of its friend list. Qualitatively perusing the average CSNP's friend list, it doesn't take long to notice the abundance of unrelated CSNPs and spam profiles within the list. Spam profiles are CSNPs disguised as personal profiles. Telling spam profiles apart from personal profiles is a topic onto its own and thus beyond the scope of this chapter. Suffice to say that common-sense observation should allow an individual with intermediate knowledge of social networks to discern between a genuine personal profile and spam profile. It is imperative to understand the importance of the need to categorize connections as valid and invalid.

Take into consideration a situation where a CSNP promoting a performing artist has one million friends. Qualitative analysis has helped us determine that 9,99,000 of those friends are other performing artists and spam profiles. Under quantitative analysis, it can be touted that this CSNP has achieved significant market reach and overall success due to the fact that its friend list contains one million friends. Qualitatively evaluating the friend list and counting only those friends that are not other CSNPs or spam profiles shows that the performing artist has a rather paltry following and not anywhere near as pervasive in the market as quantitative analysis showed. Heuristic two defines a simple way to determine whether a member of a CSNP's friend list should be considered valid for purposes of analysis.

Heuristic 2:

- Only those friends on a CSNP's friend list that are not CSNPs or spam profiles should be considered valid friends.

5.4.1.2 Connection Initiation

In addition to determining the number of valid connections in a friend list, it is helpful to examine the source of the valid connection. As mentioned earlier, there are only two ways to add friends to a friend list. A CSNP can send a friend request to a personal profile or the personal profile can send a friend request to the CSNP. Though a connection may be valid; its source, when looked at qualitatively, can yield more information as to the weight that should be given to it.

A connection initiated by a CSNP to a personal profile for purposes of product promotion shows a proactive effort on the CSNP's part to grow their customer base. On the other hand, a connection initiated by a personal profile towards a CSNP shows an unprovoked overt act by an individual in support of the product associated with the CSNP. This is a clear indication of the individual's true support of the CSNP's product.

Heuristic 3:

- A connection request initiated by a personal profile to a CSNP demonstrates an individual's proactive support for the product associated with the CSNP and thus carries more weight than a CSNP initiated connection request.

5.4.1.3 Connection Demographics

Incorporating demographic information in the analysis of personal profiles is difficult to achieve due to the absence of information verification mechanisms. The owner of a personal profile could claim to be a 50-year-old Asian female in Los Angeles and could, in reality, be a 20-year-old white male in New York. This makes it very difficult to apply any sort of demographic based heuristic to CSNP analysis. If it were possible to verify the user information pertaining to each friend in a CSNP's friend list; it would be prudent to introduce a heuristic to our quantitative analysis. This heuristic would give greater weight to those profiles that fall within the target demographic, as defined by the commercial enterprise behind the CSNP. Demographic information can include age, gender, location, musical tastes, buying patterns, education, yearly income, and more.

In order to fully understand the importance of incorporating a demographic-based heuristic into CSNP analysis, picture a pop-rock band with a CSNP. Their tour manager is trying to plot their tour route. Without the aid of a demographic-based heuristic, CSNP analysis serves no purpose for his situation. If it was possible to verify even a fraction of the demographic information in a profile, the tour manager would be able to analyze his band's friend list to a level of detail that would

be of great use to him. For instance, if the geographic location of each friend in the band's friend list became verifiable; the tour manger would be able to create a map overlay and visualize the areas on a map with the highest concentration of the band's friends. This would help him plan a tour route that stopped at locations with high concentrations of the band's friends; thus increasing the chances for a good turnout at each performance.

In the event that further research yields a reliable means for demographic information verification pertaining to standard personal profiles, heuristic 4 will be introduced.

Heuristic 4:

- If a friend's demographic information conforms to a set of specified desirable parameters it, indicates the friend will be more likely to actively support the CSNP's product.

5.4.1.4 Friend Stacking

Friend stacking is a deceptive technique by which CSNPs gain a large number of friends even before they have officially announced the profile to the public [11]. One of the approaches often seen begins with a seemingly legitimate profile for a charitable cause; such as the search for a missing child. In reality it's actually a commercial enterprise behind the site. Automated scripts begin sending thousands of friend requests as part of a coordinated social network spam campaign. The good hearted nature of many social network users causes them to accept the friend request in order to help the charitable effort.

Once enough of the requests have been accepted, the commercial enterprise changes the content of the profile and turns it into a CSNP. Now they have a CSNP that is launched with several thousand friends already present in the CSNP's friend list. This creates the appearance of a sweeping phenomenon that people cannot wait to be a part of. After the CSNP's launch the automated scripts continue to send friend requests; this time without the pretext of the charitable cause. The fact that the friendees see a request from a CSNP with several thousand friends makes it more likely that they will accept the friend request as opposed to seeing a request from a CSNP with a very low number of friends. Friend stacking takes advantage of mob mentality by creating the impression that their CSNP's product already has a devoted following. Taking this into account, it is reasonable to infer that the results of any analysis performed on a CSNP that has benefited from friend stacking is, at minimum, skewed.

Heuristic 5:

- If friend stacking was used at anytime during the lifetime of a CSNP, any conclusions obtained through analysis will be skewed; unless it is possible to identify and remove all connections made through friend stacking.

5.4.1.5 Div Overlaying

A less complicated alternative to friend stacking is a technique known as Div Overlaying. Simply put, an HTML div element with a z-index greater than that of any other element in the profile is positioned over the CSNP's friend list. The content of this div is indiscernible from the real friend list except for the fact that the text indicating the number of friends has been modified [11]. The aim is to deceive the average user into thinking that the CSNP has a greater number of friends than it actually has. Since the content of the div overlay is nothing more than standard HTML, the CSNP's administrator can choose to display any number they wish. They can even periodically increment the number so as to give the illusion of legitimate increases in friend list size.

A straight forward way to determine whether a CSNP is employing Div Overlaying is to select the option to view the complete list of the CSNP's friends. This will show the true number of friends as the use of Div Overlaying is, more often than not, limited to the main profile page. Div overlaying is not a technique that can be used on all social networks; though it is one of the most popular hacks used on MySpace personal and specialized profiles alike. The use of Div Overlaying extends far beyond merely changing the apparent size of a friend list. Div Overlaying has been used to replace entire profile main images. Examples can be found on every other CSNP throughout MySpace and other popular social networks.

Heuristic 6:

- When conducting analysis that takes into account the size of a friend list, a check for the use of Div Overlaying must be performed in order to insure the accuracy of the friend list size.

5.4.2 Interactions

As with connections, quantitative analysis of interactions based simply on raw numbers can be very misleading. Qualitative factors must be weighed-in if accurate analysis is to be achieved; most notably the type, source, and content of the interactions.

5.4.2.1 Interaction Type, Source, and Content

Interaction types can be wall posts, comments, picture ratings, bulletin board replies, e-props, and any other interaction defined within a social network. The significance, or weight, given to each type of interaction will be determined by nature of the analysis being conducted. For instance, if analysis is being performed in hopes of determining the impact a particular bulletin board entry has had on the overall CSNP then, emphasis will be placed on bulletin board replies.

The source of an interaction is determined by identifying the origin of the interaction. For an unrelated CSNP to personal profile interaction the source is the CSNP, as the interaction was initiated by the CSNP. When it comes to related interactions, the source of each subsequent interaction is the same as the source of the first interaction of the series of related interactions.

Content is a factor that is difficult to evaluate yet very valuable for analysis. The content of an interaction needs to be semantically analyzed in order to assign a proper weight to it. These three key factors should be considered concurrently in order to yield the most accurate analysis possible. To demonstrate this, consider the following two unrelated interactions and the analysis that follows.

Interaction *A*

Social Network	Facebook
Type	Wall Post on Bob's Personal Profile
Source	Acme Inc.'s CSNP
Content	"Hi! Do you like our product?"

Interaction *B*

Social Network	Facebook
Type	Wall Post on Acme Inc.'s CSNP
Source	Jane's Personal Profile
Content	"I enjoy your product! I use it all the time!"

Note: Both interactions are unrelated to any other interaction

Under strict quantitative analysis, interactions *A* and *B* are identical. If qualitative analysis of the three key factors mentioned is incorporated then it is clear that the resulting analysis will indicate a significant difference between the two interactions.

A cursory glance reveals two noticeable differences, the first being that the sources for the interactions are different. Interaction *A* is CSNP to personal profile while interaction *B* is personal profile to CSNP. Secondly, the content is very different. The content of interaction *A* appears to be an attempt by Acme Inc. to solicit feedback from Bob concerning their product and interaction *B* is apparently an unsolicited compliment from Jane to the Acme Inc. regarding their product. As interaction *A* is unrelated to any other interaction, was initiated by the CSNP, and has content intended to trigger one or more related interactions, it is not a genuine social interaction. Interactions such as interaction *A* are said to be part of interaction phishing.

Interaction phishing is an attempt by a CSNP to elicit personal profile to CSNP interactions for the sole purpose of increasing its overall and personal profile to CSNP interaction counts by baiting a personal profile into initiating social interactions. Interaction phishing may occur via CSNP initiated connection requests or CSNP initiated interactions. A personal profile to CSNP interaction that was triggered by a CSNP to personal profile connection request or a CSNP to personal profile interaction is considered to be a result of interaction phishing.

Heuristic 7:

- Interactions resulting from interaction phishing are not genuine social interactions

Unlike the case with interaction A , the source and content of interaction B leads one to conclude that interaction B was an unsolicited attempt by a personal profile to engage a CSNP in social interaction. As interaction B is unrelated to any other interaction, is originated from a personal profile, and its content is clearly indicative of the personal profile's support of the CSNP's product, then it is considered a genuine social interaction and should be weighed appropriately during analysis.

Heuristic 8:

- A personal profile to CSNP interaction between personal profile φ and CSNP γ that is unrelated to any other interaction or related to a series of interactions whose initial source was φ is a genuine social interaction.

Heuristic 9:

- A CSNP to personal profile interaction between personal profile φ and CSNP γ that is either related to a series of interactions whose initial source was φ or initiated by γ as a response to a connection request initiated by φ is a genuine social interaction.

5.4.2.2 Interaction Based Spam

Considerations pertaining to the content of an interaction should be made to account for the spam epidemic that plagues social networks [12]. Social network spam can have many forms; some of which are analogous to those found in email spam [13]. Social network spam includes offers from supposed foreign politicians, real estate deals that are too good to pass up, prescription drugs at prices too low to believe, and more. In addition to these well known forms of spam, social networks fall victim to social network specific spam such as interaction phishing, ballot stuffing, and friend stacking just to name a few. Interaction based spam can greatly skew any kind of quantitative analysis results unless the proper heuristics are applied. The topic of social network spam deserves in depth consideration far beyond the scope of this chapter. This section is limited to the introduction of a couple of heuristics that may help negate or at least lessen the effect that interaction based spam has on the analysis of CSNPs.

Heuristic 10:

- Any unrelated interaction whose sole purpose is to solicit the interaction recipient is considered social network spam.

It is important to note that interaction based spam can originate from either a CSNP or a personal profile. Personal profiles are capable of creating interaction based spam with things such as birthday announcements, links to surveys, holiday tidings, joke links, and countless other things. CSNPs create interaction based spam with a much more specific purpose, to solicit the interaction recipient. Evidence of CSNPs using interaction based spam can be seen throughout each of their friend's profiles. CSNPs are known for initiating random unrelated interactions at specific times, such as product release dates, on their friends' profiles. Some of the more notorious producers of interaction based spam are music CSNPs.

Music CSNPs will often use interactions to announce every single stop on their tour. If a band with a CSNP is playing in New York on January 1, 2010 then they will initiate interactions with as many friends as possible; soliciting their attendance to the performance. Tactics such as this create interaction based spam equivalent to the number of stops on the tour times the number of friends in the CSNP's friend list. As music CSNPs are known to use friend stacking [11] and other techniques to artificially raise the number of friends on their friends list, this can create millions of instances of interaction based spam.

Heuristic 11:

- Any interactions deemed to be social interaction based spam are not genuine social interactions and should be treated accordingly during interaction analysis.

In determining if an interaction is spam it may be useful to apply traditional spam detection heuristics found in email and social spam detection systems [14]. The principle is the same; analyze the content of the interaction and based on the results and determine if it's spam or not. Additionally, it will be necessary to apply social network specific heuristics to obtain a better estimate of the amount of spam in the set of a profile's total interactions.

Heuristic 12:

- When analyzing interaction statistics, generic and social network specific spam heuristics must be developed and applied and any interaction deemed to be spam must be removed from consideration.

As stated earlier, social network spam and the topics it encompasses, such as social network spam detection, deserve separate and in depth attention beyond the scope of this chapter. Suffice to say, a good foundation for the development and application of social network spam detection and management can be found using heuristics 10 and 11.

5.4.2.3 Interaction Frequency and Timing

Two attributes of social network interactions rarely taken into account during analysis are frequency and timing of interactions. That is, when a set of interactions occur and how often they occur over a set period of time. When it comes to CSNPs the frequency and timing of interactions are valuable pieces of information. They help commercial enterprises directly correlate past events to customer reaction and anticipate customer responses to future events. In the case of band's CSNPs, it is useful to see what kind of a reaction the band receives from their social network friends who attended their performance as well as gauge how many friends will be attending the performance to begin with.

For instance, the week leading up to a performance in Chicago a band receives 1,000 genuine interactions and valid connection requests initiated by personal profiles in the Chicago area. This can be taken as an indication that the band can expect a great turnout for their performance in Chicago. The same could not be said if the

band received the same 1,000 interactions over the course of the 6 months prior to the show due to the uncertainty large spans of time cause in the entertainment industry. In the week following the band's performance in Chicago, their CSNP receives 1,000 more genuine interactions and valid connection requests from personal profiles in the Chicago area. This can be taken as a clear indication that their performance in Chicago was a great success. On the other hand, if the band had received only 100 genuine interactions and valid connection requests it could be said, with considerably less certainty, that their performance was not as successful as it could have been. In many cases, increases in interaction frequency around a specific time can help tie the jump to a specific event.; while a lack of increases in interaction frequency around a specific time are not as easily tied to a specific event. This is simply because the nature of social network interactions is that users will not always interact. This means that the lack of interaction may be attributed typical user patters as opposed to the lack of interaction being caused by a specific event, such as a performance or product release.

Heuristic 13:

- If there is an increase in the number of genuine personal profile to CSNP interactions and valid personal profile to CSNP connection requests received during the period of time immediately preceding and or following an event, it can be reasonably inferred that the event in question is in part responsible for the increase.

Having interaction frequency and timing does not limit one to tying increases in interactions to events. Knowing the frequency and timing of interactions can also help analysts determine how fresh the success of a CSNP is. Take into consideration a case where as CSNP has 1,00,000 valid connections and over its lifetime has participated in 2,00,000 genuine social interactions. At first, it seems as though the CSNP is currently enjoying the heights of success. Now consider additional information has come to light. The latest interaction occurred 6 months ago and in the 2 years preceding that latest interaction, the interaction frequency was six interactions per week.

Heuristic 14:

- When assessing a CSNP's current level of success based on genuine interactions and valid connections, it is necessary to evaluate the frequency of each as compared to the average frequency over the lifetime of the CSNP. If the recent frequency is slightly less than, equal to, or greater than the average overall frequency then the CSNP's success can be considered current.

With this information and the application of heuristic 14, it can be ascertained that in the past two and a half years there have only been 624 interactions that involved this CSNP. One can now infer that the bulk of this CSNP's success happened over two and a half years ago; clearly a sign that this CSNP's best days are behind it. A far cry from the results obtained without the use of heuristic 14.

5.4.2.4 Interaction Uniqueness

Uniqueness of interactions is a brief yet important topic. When analyzing a CSNP one must look at the breakdown of the personal profiles interacting with it. If it is found that any personal profiles take part in a much greater percentage of interactions than average, the value of those interactions must be weighed accordingly. A case might exist where a CSNP has 1,000 genuine interactions but, all of those interactions come from a single source. That should not be evaluated the same as a CSNP with 1,000 interactions that each come from a unique source.

Heuristic 15:

- For every personal profile α in the set of all personal profiles that have been identified as a source of a genuine interaction between itself and the CSNP being analyzed, the number of interactions per α should be within one standard deviation of the total number of interactions divided by the number of unique sources. Otherwise it is considered an outlier and removed from the analytical process.

5.5 Technical Notes

In obtaining the necessary quantitative and qualitative data for this research various approaches were undertaken. In order to develop the heuristics introduced in this chapter it was necessary to have the ability to crawl numerous social network profiles and extract pertinent information from each one. This task was accomplished thanks to tools and methods introduced to the authors by their industrial partners at IBM Almaden Research Center and Barcelona Supercomputing Center (BSC).

Two of the more vital tools were a social network crawler and profile information extractor. Both were developed using IBM and BSC methods; such as SystemT [15] and MapReduce [16]. The social network crawler would retrieve a random sampling of social network profiles and store the results in a series of data files containing the raw HTML of each profile. The profile information extractor would then process these files and using a set of regular expressions, return an organized dataset that allowed us study the results and from there, create the heuristics introduced in this chapter.

In addition, various other scripts in combination with manual methods were used to achieve as accurate a picture as possible of the intricacies behind CSNPs. This chapter will attempt to paint for the reader that very same picture.

5.6 Summary

The furious pace at which social networks and CSNPs are evolving makes it imperative to continually monitor industry trends. The heuristics introduced here are presented in a generic manner to make it easy to adapt them to the inevitable changes

in the world of social networks. Much work is left to be done in order to fully understand the impact of CSNPs. A fusion of marketing, sociology, and computer science research is necessary in order to paint a detailed picture of the effects CSNPs have on industry and society.

Acknowledgement This research was supported in part by NSF grants IIS-0837716, CNS-0821345, HRD-0833093, IIP-0829576, IIP-0931517, DGE-0549489, IIS-0957394, IIS-0847680, CNS-0837556 and CNS-0426125. IBM Almaden Research Center and Barcelona Supercomputing Center have contributed to the methodology used and to the first author's training under the NSF PIRE award 0730065.

References

1. "Signup For MySpace, Now with Free Music." Signup for Myspace, Now with Free Music. MySpace.com. Web. 18 Mar. 2010. <http://signups.myspace.com/index.cfm?fuseaction=signup>
2. "Fansite – " Wikipedia, the Free Encyclopedia. Wikimedia Foundation, Inc. Web. 18 Mar. 2010. <http://en.wikipedia.org/wiki/Fanpage>
3. "The Drive Home Only Adds REAL People => on MySpace Music – Free Streaming MP3s, Pictures & Music Downloads." The Drive Home Only Adds REAL People => on MySpace Music – Free Streaming MP3s, Pictures & Music Downloads. Drive Home Music Inc. Web. 18 Mar. 2010. <http://www.myspace.com/drivehomemusic>
4. "Twilight Saga New Moon Official MySpace Profile." Twilight Saga New Moon Official Myspace Profile. Summit Entertainment LLC. Web. 18 Mar. 2010. <http://www.myspace.com/twilight>
5. "MySpace – It's Always Sunny in Philadelphia – Only on FX!" It's Always Sunny in Philadelphia – Only on FX! FX Networks LLC. Web. 18 Mar. 2010. <http://www.myspace.com/sunnyfx>
6. "Dane Cook on MySpace Comedy – Comic Clips, Funny Videos & Jokes." Dane Cook on MySpace Comedy. Web. 18 Mar. 2010. <http://www.myspace.com/danecook>
7. Vans Warped Tour 2010. Vans. Web. 18 Mar. 2010. <http://www.myspace.com/warpedtour>
8. "Social Relation" Wikipedia, the Free Encyclopedia. Wikimedia Foundation, Inc. Web. 18 Mar. 2010. http://en.wikipedia.org/wiki/Social_interaction
9. Melendez, Lester A. Crawl of 1000 MySpace Profiles to Gather Interaction Statistics. 10 Mar. 2010. Raw data. Florida International University, Miami
10. "Two Thirds of Businesses Fear That Social Networking Endangers Corporate Security, Sophos Research Reveals." Antivirus | Security Software | Data Protection | Encryption Software for Businesses – Sophos. Sophos Plc. Web. 18 Mar. 2010. <http://www.sophos.com/pressoffice/news/articles/2009/04/social-networking.html>
11. "Cori Yarckin on MySpace Music – Free Streaming MP3s, Pictures & Music Downloads." Cori Yarckin on MySpace Music. See Why Records. Web. 18 Mar. 2010. <http://www.myspace.com/coriyarckin>
12. "Malware and Spam Rise 70% on Social Networks, Security Report Reveals." Antivirus | Security Software | Data Protection | Encryption Software for Businesses – Sophos. Sophos Plc. Web. 18 Mar. 2010. <http://www.sophos.com/pressoffice/news/articles/2010/02/security-report-2010.html>
13. "LOVE SHE WROTE on MySpace Music – Free Streaming MP3s, Pictures & Music Downloads." Love She Wrote on MySpace Music. Web. 18 Mar. 2010. <http://www.myspace.com/loveshewrote>
14. Benjamin, M., Cattuto, C., and Menczer, F. 2009. Social Spam Detection. Proceedings of the 5th International Workshop on Adversarial Information Retrieval on the Web. ACM International Conference Proceeding Series, Spain, Madrid. New York: ACM Press, 41–48

15. Krishnamurthy, R., Li, Y., Raghavan, S., Reiss, F., Vaithyanathan, S., and Zhu, H. 2009. SystemT: A System for Declarative Information Extraction. *SIGMOD Record*, 37:4, 7–13
16. Dean, J. and Ghemawat, S. 2008. MapReduce: Simplified Data Processing on Large Clusters. *Communications of the ACM Archive*, 51:1, 107–113