Mark Rogers

received his MA (First Class) in English Literature from Oxford in 1983. He co-founded www.bbc.co.uk as commissioning editor in 1997 and has since worked as a product manager for Amazon Anywhere and a number of start-ups in Sweden, US and UK. He is CEO of Sentinel Projects Ltd.

Clovis Chapman

received an MSc in Data Communications, Networks and Distributed Systems and a PhD in 2009 from the department of Computer Science at University College London. He has since been working on large-scale distributed system and data analytics, focusing in this context on conversational analytics for social media platforms such as Facebook and Twitter.

Vasileios Giotsas

received the MSc in Data Communications, Networks and Distributed Systems from University College London (UCL), UK in 2008. Currently he is a PhD candidate at the Network Research Group, Department of Computer Science of UCL. His research interests include analysis and modelling of the complex networks with focus on networks economics, topology dynamics and Internet routing.

Keywords: Facebook, marketing, brand, diffusion, social media

Mark Rogers
Sentinel Projects Limited
155 Kennington Park
Road London SE11 4JJ
UK
E-mail: mark.rogers@sentinel-projects.com

Papers

Measuring the diffusion of marketing messages across a social network

Mark Rogers, Clovis Chapman and Vasileios Giotsas

Received (in revised form): 27th June 2012

Abstract

The last few years have seen significant investment in social media as an advertising, marketing and customer outreach opportunity. In the US alone, in 2010, almost \$1.7 bn was spent by advertisers on social media marketing, with 53 per cent specifically allocated to Facebook¹. Due to the explicit links that users maintain with each other, social media platforms are perceived as a highly suited environment for network-based marketing: *word-of-mouth marketing, diffusion of innovation, or buzz and viral marketing*² all aim to take advantage of the relationships between users to facilitate the spread of awareness or adoption. In order to predetermine the effectiveness of such campaigns, it is important to be able to estimate potential return on investment. In particular, the ability to model existing networks, track the propagation of marketing messages and estimate customer exposures and impressions are essential for this purpose. A wide range of techniques to measure notions such as user engagement on such platforms have been developed and there also exists a significant amount of research on modelling contagion and diffusion in network-based environments that can be exploited to generally refine an overall marketing strategy. However, the structure and properties of different social media platforms introduce various constraints on both the means via which data propagate and the visibility of content and nodes, constraints that must be taken into account when modelling or measuring the impact of social media campaigns. Perfect information about exposures within a given graph to a given message will not be available and as such it is important to investigate and define methodologies for diffusion monitoring that are suited to specific platforms.

Journal of Direct, Data and Digital Marketing Practice (2012) **14**, 97–130. doi:10.1057/dddmp.2012.25

Introduction

Purpose of this document

The overall objective of this document is to identify which social media platform is most suited for the purpose of network-based

marketing, and the methodologies that can be exploited to correctly monitor the impact of various marketing strategies. In particular, this document will aim to:

- Define key characteristics of social media platforms that have a potential impact on the marketing process.
- Survey existing methodological research literature and tools that allow us to reliably track the impact of this process.
- Detail an approach that enables us to test and monitor selected methodologies on a selected platform.
- Identify the nature of the data and the size of the samples that should be collected in order to reliably estimate the impact of a marketing strategy.

Scope and goals

Scope

As the field of network-based marketing is vast, and social media platforms are numerous, this document will focus on Facebook as the largest social media platform at time of writing. Facebook has an ecosystem that is relatively diverse: it maintains social links between users and it provides tools and options for communities and groups to form and interact in the context of particular topics/brands, for brands to communicate with users, and users to control public feeds/streams of information. Privacy constraints, however, have imposed significant restrictions on the ability to crawl a user's social graph, and this document will aim to examine if these limitations can be overcome to provide significant insight into message propagation among users.

In addition this document will focus on the process of monitoring message propagation and diffusion on the Facebook social graph, as opposed to examining marketing strategies. The overarching goal is to define requirements for a monitoring framework specific to Facebook. These requirements will be based on the conclusions of the survey.

Structure of the document

This document is structured as follows: in the next section, we provide background information on the Facebook platform, detailing its key characteristics and operation, and identify monitorable attributes that can serve as a basis for defining suited metrics. After that, we survey existing state of the art and research in the general domain of network-based marketing. This spans a number of fields, such as economics, psychology and computer science, and we attempt to identify key contributions that are directly relevant to the process of social media marketing and message diffusion. In the subsequent section, we detail potential approaches for the purpose of monitoring message propagation in Facebook, identify the limitations of such approaches and estimate the data sampling required to validate the approach. Finally we conclude with an overview of the requirements for the development of a Facebook-based monitoring framework.

The Facebook ecosystem

Background: Facebook

With over 900 million active users, Facebook is currently the largest online social network;³ ranking second in global traffic, closely behind Google, and widely deployed globally with translations in 70 official languages and application development taking place in over 190 countries. The Facebook ecosystem can be divided into a number of subsets: the friendship network constitutes the core of Facebook's offering, allowing individual users to create self-descriptive profiles that include links to their 'friends', links that must be confirmed by both sides, ensuring a network of reciprocated ties. This core network contains a significant amount of demographic data and detailed personal information (status, photos, relationships, interests, current location, etc) published through status updates (a *wall*), configuration settings and likes. A focal point of interaction of a user with Facebook is the news feed, which shows updates of friends and their activities on Facebook.

Applications and Fan pages

The Facebook platform environment provides the means for developers to create custom applications through the use of the *graph API*. This API, central to Facebook's marketing strategy, allows third parties to integrate applications into the core facebook experience, requesting specific permissions to obtain from users' potentially private data relevant to the application domain (eg friend list, access to wall posts, demographics, likes, etc). By way of example, the 80 applications produced by Zynga, Facebook's largest developer (Farmville, and Mafia wars), have accumulated over 250 million users, more than half of the Facebook worldwide count.⁴ Generally content and interactions produced via the application may be posted on a user's update stream, potentially emerging on friends' news stream.

Facebook fan pages provide a context through which users can interact with respect to a product/brand or topic/cause of choice. Fan pages provide discussion walls or forums in addition to topic-specific information and have provided a popular channel for the launch of marketing campaigns on Facebook, as a means of highlighting specific products and services and nurturing brand communities. Facebook statistics⁵ show over 3 million active fan pages, with the top 20 having individually over 20 million fans. With the majority of these pages being public (or accessible with appropriate credentials), a significant amount of data can be harvested giving us key insights into community interactions. Similarly to applications, interactions with a fan page of a friend may appear in a user's news stream, resulting in third party exposures.

Information diffusion on Facebook

The news feed

As previously stated, the primary source of interaction with Facebook is the *news feed*. Estimates show that nearly 80 per cent of interactions with Facebook take place via the news feed, making it a highly

important mechanism for diffusion of activities and broadcast of information.

Information sharing via the news feed can be both active (user posting on a wall) or passive: actions of users, such as interactions with other users, applications or fan pages, will be broadcast to their entire network of friends. The actions that appear on a user's news feed are aggregated and ranked according to an algorithm based on social and content-based features.

Users themselves do exert a certain level of control over the content that appears, including the ability to comment, like or even hide the updates; the latter having a potential effect on the likelihood of appearance of future similar updates.

The EdgeRank algorithm

The algorithm, named EdgeRank,⁶ ensures that relevant content appears in a user's news feed by taking into account affinities between producers of content and their consumers, the timeliness of the action and the perceived value of the content. More specifically, news feed items are considered objects and interactions with an object results in the creation of an edge, which includes actions such as comments and tags. As such the algorithm can be briefly described as follows:

$$\sum_{\text{edges } e} u_e w_e d_e \quad (1)$$

where u_e represents the affinity score between viewer and creator, determined from historical interactions between the users (message exchange, profile views, etc), w_e represents the weight of the edge type (likes may rank differently to comments), and d_e represents the time decay factor.

The affinity of relationships has a critical impact on the diffusion process

Crucially, while this algorithm highlights the notion that network relationships have a fundamental impact on the diffusion of content on the platform, the structure of the network provides only a partial insight into the diffusion process. An accurate estimation of the diffusion process can only be obtained through an understanding of the relationship between users as represented by their affinity (where for example updates by old acquaintances will have a lesser probability of appearing on the news feed than those of close friends). Estimating news feed impressions may require examining a wider range of user characteristics (demographics, etc) where observations of previous interactions are not available.

Diffusion via social applications

Viral product features

Alongside the active propagation of messages via user to user interactions, and passive diffusion via the news feed, applications built on the Facebook graph may implement various methodologies to actively reach and encourage engagement with new users. Such viral product features may include forms of *personalised referrals*, such as user to user invitations. In this instance, users can select friends or

contacts from a list and invite them to join or install the application. Alternatively specific features of the application itself may involve targeting friends and contacts as part of the service itself: Branchout, a professional networking service, encourages users to answer questions about their friends and posts that information on their wall. On the other hand, a Snickers-related application, built to promote the Snickers Peanut Butter Square, enables users to upgrade 'likes' to 'loves'. The use of such an application results in a marketing message being posted in a user profile, designed to result in potential third party impressions and further application installs.

Advertising on the Facebook platform

Targeted advertising

We next consider targeted advertising on the Facebook platform as a mechanism for reaching particular demographics of users. Facebook allows the creation of custom advertisement messages through which one can target a specific message and link to users according to location, demographics (including sex, age, language, relationship status and education) and interests. The interest feature in particular rely on the past 'likes' and other affinities of users to effectively target communities of users — providing an insight into the reach of specific advertisements.

Facebook's advertisement platform also aims to take advantage of interlinks between users in the form of *sponsored stories*. Sponsored stories allow businesses to surface word-of-mouth recommendations that exist in the Facebook news feed — effectively amplifying news feed items. Page likes, application use or link shares can effectively be captured in the form of an ad. Such stories are restricted according to privacy settings and other eligibility criteria.

Viral feature space

Viral features shape a product's use in relation to other customers

Aral and Walker⁷ examine the effectiveness of various viral features on social networking platforms. While their primary focus is on comparing automated broadcasts to personalized invites, they attempt to classify viral features of products according to the level of active involvement of the user required and the personalization levels of the feature itself. A product's viral features shape and constrain a product's use in relation to other consumers. As such, it can be hypothesized that these features will have a significant impact on the reach and diffusion of a specific marketing message, alongside network structure, influence, demographics and interests.

We rely on a similar approach to classify the diffusion features of Facebook, as illustrated in Figure 1. Alongside the features discussed above, such as news feed broadcasting, sponsored stories and personalized invitations for application use, we can also consider additional features: user-generated content whether generated via an application (eg photos, videos) or externally produced and shared with other users via some form of hypertext embedding (eg YouTube videos). Direct user to user messaging, whether private or via a wall

Figure 1: Facebook viral feature space

The invitation fatigue effect

post, may also figure as a diffusion mechanism, with the potential sharing of links, product information or brand mentions.

To demonstrate the effectiveness of various viral features, the authors rely on a custom application on Facebook with the ability to assign, or disable, different feature sets for different subgroups of users. The paper compares the relative effectiveness of invitations sent upon installation of the application, against the passive broadcasting of actions on the news feed. This is achieved using hazard modelling, the standard technique for assessing contagion in economics, marketing and sociology, which involves representing the hazard of adoption of individual i at time t as a function of individual characteristics and social influence. The authors conclude that passive broadcasting resulted in a 246 per cent increase in the rate of application of adoption whereas adding application invitations only resulted in an additional 98 per cent increase. While no concrete explanations are given for this discrepancy, we may hypothesize that repeated broadcasts of similar actions such as invitations from a wide range of applications may reduce the overall user experience and introduce a ‘invitation fatigue’ effect. Users may respond better to observing multiple friends using an application via their news feed rather than to an explicit invitation to use an unknown application, implying that a certain threshold of adoption may have to be reached before contagion effects can take effect, as will be discussed later. As a whole, this work provides an interesting methodology for evaluating and comparing various viral feature sets at various levels of granularity.

Facebook graph objects

The Graph API

In order to derive key metrics that can be used to measure the propagation of data on Facebook, we identify in this section relevant data that may be gathered from the Facebook Graph API. The Graph API presents a view of the Facebook social graph, uniformly representing objects in the graph (eg people, photos, events and pages)

and the connections between them (eg friend relationships and shared content). Though all public information can be gathered via this API, getting additional information requires explicit permission. We distinguish here parameters that are made available publicly, and those that can be only be obtained using specific permissions. The access levels are categorized as follows and detailed in Table 1:

- *Public [P]*: Accessible without any explicit permission
- *Permission Required [PR]*: The user must explicitly authorize access to this information, primarily through a third party application.
- *Administrative Access Required [AR]*: This information can only be accessed by the owner of the object (fan page, application, etc)

Access levels of Facebook attributes

The table shows that only limited subsets of the Facebook Graph API are publicly accessible. Other aspects may require explicit permission, and developers may request such permissions should the nature of the application justify the use of these data.

Table 1: Key attributes of graph objects

Category	Attribute	Detail	Access level
User	Name		[P]
	Gender		[PR]
	Location		[PR]
	Work		[PR]
	Relationship status		[PR]
	External website		[PR]
	Friend list		[PR]
	Likes	Pages liked	[PR]
	Wall posts	Content/Feedback/Likes/ Created time/author	[PR][P]
	Comments	Content/Likes/author/time	[PR][P]
	Movies/music/books/...		[PR]
	Photo tags/albums		[PR]
	Video tags/uploads		[PR]
Application	Installs		[AR]
	Active users		[AR]
	Stream published stories		[AR]
	Status updates		[AR]
	Links posted		[AR]
	Photos created		[AR]
	Application user demographics		[AR]
	Content feedback	News feed comments, news feed likes	[AR]
	Hides		[AR]
	Requests sent	Accepted/ignored/etc	[AR]
Fan pages	Fans (likes)		[P]
	Fan user list		[P]
	Wall posts	Content/Feedback/Likes/ Created time/author	[P]
	Wall comments	Content/Feedback/Likes/ Created time/author	[P]
	Impressions	News feed/Logged in users	[AR]
	External referrers	Feedback/Comments	[AR]
	Demographics	Location/Age/Male Female ratio	[AR]

Content engagement metrics

From this initial set of measurable attributes, a number of content engagement metrics can be derived to monitor the relative success of a Facebook campaign. Fan page analytics for example may rely on fan growth rate over specific periods as an awareness metric, administrator to user content ratios may serve to measure user participation, and response rates, in the form of 'likes', comments or news feed impressions may serve as an approximation of reach. Insight into the diffusion process of messages may be obtained through extrapolation of these data, building for example pictures of users, behaviour and interactions, which serve to complement network information. We further discuss the relationship between such metrics and return on investment in the section 'Challenges and research opportunities'.

We examine in the following section the methodologies that can be exploited to effectively overcome or compensate for these privacy constraints, and the analysis that can be performed on these metrics.

Related work and state of the art

Survey of existing literature

Related literature

To guide the overall literature review, we refer to the survey of network-based marketing methodologies conducted by Hill *et al.*² The authors provide an overview of key strategies for identifying likely adopters through consumer networks. Their research provides a detailed overview of existing work in this domain, and highlights the challenges and issues encountered when dealing with econometric models, network classification models, surveys, designed experiments, diffusion models and collaborative filtering systems. Briefly, econometric models provide empirical estimations of economic relationships through statistical methods. Network classification models aim to quantify interest between entities in a network (eg Google's PageRank algorithm). Traditional surveys enable consumer habits about word of mouth behaviour to be collected. Designed experiments provide controlled settings through which one can observe interactions between subjects. Diffusion models provide tools for assessing the likely rate of diffusion of a technology or product in networks. Finally, collaborative filtering systems make personalized recommendations to individuals based on demographic content and link data.

Moreover they propose empirical evidence that statistical models built from a combination of large amounts of geographic, demographic and prior purchase data are significantly improved through inclusion of network link information. Such models can be used to more effectively target potential consumers when relying on network-based marketing methodologies.

This survey allows us to broadly identify the approaches most likely to be suited to Facebook, and the metrics most relevant in the context of each approach. However, the focus of the survey does not take into account the specific nature of social media platforms in general and Facebook in particular, which may differ from 'real-world' social

environments and may present other characteristics and data sets relevant to this work, such as a wealth of text-based conversations. As such, it is necessary for us to broaden the literature review to encapsulate additional research in this domain. We categorize existing work according to the following research areas: network diffusion models (section ‘Information diffusion’), collaborative filtering and recommender systems (section ‘Recommender systems’) and semiotic and linguistic analysis (section ‘Semiotic and semantic analysis’) and provide for each category a brief overview, an analysis of work specifically related to our problem domain and relevant to the context of this document and detail the potential approaches suited for Facebook and their potential caveats.

Theory of information diffusion

Information diffusion

Diffusion theory provides tools to assess the likely rate of diffusion of a technology or product. Understanding the individual-level data and connections between users of a social network has become of great value to marketers due to the impact of the network structure on the diffusion of a marketing message.

We begin in this section by providing an overview of network statistics, and information diffusion and its relation to social media environments and marketing. This overview aims to provide clear definitions for notions such as diffusion and adoption and provide the necessary background for this work. We then examine related work focused specifically on social media platforms such as Facebook and Twitter, and finally identify limitations with this work and potentially applicable approaches.

Graph-theoretic definition of social networks

Statistics of networks structure. Any social network can be described as a graph $G=(V,E)$, where V is a non-empty set of *nodes* (or *actors*) and E is a set of *edges* (or *links*) that connect pairs of nodes. The actors correspond to individual persons in the social network and the links correspond to some type of social relationship between the actors (eg friendship). The number of actors $n=|V|$ is the *order* of the graph while the number of links $m=|E|$ is the *size* of the graph. The *adjacency matrix* A is a $n \times n$ matrix that indicates which links exist between the actors. If the actors v,w are connected by a directed link (from v to w), then the element $a_{v,w}$ is equal to 1. Otherwise it is equal to 0. If the edge is undirected then $a_{v,w}=a_{w,v}=1$. In a social network information may flow in a bi-directional manner (Facebook friendship links must be reciprocated), so we consider here primarily undirected graphs. However, other social networks such as Twitter can have uni-directional links, where following a user does not have to be reciprocated. In addition, there are no links that have as endpoints the same actor (self-loops), so the adjacency matrix of a social graph is symmetric and its diagonal elements $a_{i,i}$ are equal to 0.

A number of statistical measurements have been proposed to quantitatively characterize complex networks and study their topological properties. These measurements can be used to analyse and

understand the topological characteristics of social networks and validate the fidelity of synthetic models that try to reproduce their function. We describe below the most important topological metrics and analyse their physical significance. An extensive survey of such measurements is provided in Costa *et al.*⁸

The *degree* k_i of a node i is the most basic metric for the importance of a node and indicates the number of links (eg acquaintances) adjacent to it. For undirected graphs:

$$k_i = \sum_j a_{ij} = \sum_j a_{ji} \quad (2)$$

The *average degree* $\langle k \rangle$ of a network is the average of all k_i in the network:

$$\langle k \rangle = \frac{1}{N} \sum_i k_i = \frac{1}{n} \sum_{ij} a_{ij} \quad (3)$$

or

$$\langle k \rangle = 2m/n \quad (4)$$

Topological properties of networks

High average degree means that the network is densely connected. However, it is a very coarse metric and cannot be used for detailed analysis as graphs with the same average degree may have radically different topology.

The *degree distribution* $P(k)$ of a graph is the probability that a randomly selected node has degree k . If $n(k)$ is the number of node with degree k then the degree distribution can be calculated as

$$P(k) = n(k)/n \quad (5)$$

The degree distribution is a more informative characteristic, from which we can also calculate $\langle k \rangle$ as:

$$\langle k \rangle = \sum_{k=1}^{k_{max}} kP(k) \quad (6)$$

For graphs where any two nodes are connected with equal probability the degree distribution is binomial or Poisson for sufficiently large graph size. A well-known example of such graph is the Erdős-Rényi random graph model.⁹ On the contrary real graphs have long-tails and many of them follow the power-law distribution.

$$P(k) \sim k^{-\gamma} \quad (7)$$

Power-law networks

Graphs with power-law degree distribution are also called scale-free.¹⁰ The actual meaning of power-law degree distribution is that most of the nodes have relatively small degree but a very small number of nodes have disproportionally larger degree. Naturally these ‘rich’ actors function as information hubs for the poorly connected nodes.

If there is a positive correlation between the degrees of the connected actors, the network is *assortative* while if there is a negative correlation the network is *disassortative*. Thus, in assortative networks the actors prefer to connect to other actors of similar degree while in disassortative networks the actors seek connections with nodes of dissimilar degree.

The *clustering coefficient* c provides information on the density of connections in the neighbour of an actor. Clustering indicates the path diversity around an actor and therefore when some information reaches an actor with high clustering, the information will spread with high probability throughout the cluster. If two neighbours of a node are connected then a triangle (or 3-cycle) is formed between these nodes. The maximum number of triangles for a given node with degree k_i is $k_i(k_i-1)/2$. Therefore, the clustering coefficient can be expressed as fraction of the average number of triangles over the total number of possible triangles:

$$c = \frac{k_t}{k_i(k_i-1)/2} \quad (8)$$

where k_t is the number of triangles that the nodes shares.

Shortest path distribution: The *shortest path length distribution* $P(h)$ is defined as the probability distribution that two random actors are at minimum distance of h hops from each other. A summary statistic is the *average shortest path*

$$\langle h \rangle = \sum_{h=1}^{h_{max}} hP(h) \quad (9)$$

where h_{max} is called the *diameter* of the network.

For a graph with fixed average degree, if the average shortest path grows logarithmically or even slower with the growth of the graph order, then the graph exhibits the small-world property.¹¹ Graphs with power-law degree distribution (see equation (7)) with exponent $\gamma=3$ have mean diameter asymptotically $\log n / \log \log n$ ¹² while for exponent $2 < \gamma < 3$, $\langle h \rangle \sim \log \log n$ and for $\gamma > 3$, $\langle h \rangle \sim \log n$. The small-world property affects many basic properties of the graph, such as the spread of information or epidemics.

Betweenness centrality: The importance of an actor or a link in a graph is usually defined by the number of shortest paths in which this actor (link) participates. When an actor (link) participates in many

shortest paths, then this node is said to be closer to the centre of the graph. Betweenness B_v is the measurement that quantifies the centrality of an actor:

$$B_v = \sum_{ij} \frac{\sigma_{ij}(v)}{\sigma_{ij}} \quad (10)$$

where σ_{ij} is the total number of shortest paths between actors i, j and $\sigma_{ij}(v)$ is the number of shortest paths between i, j that pass through actor (link) v . To normalize the actor betweenness in order to compare different graphs, we divide B_v by $(n-1)(n-2)/2$, which is the maximum possible value of actor betweenness in a graph.¹³

Centrality is important in analysing information flows in a network, discovering which links can serve as a bridge between different clusters.

Coreness: The k -core of a network can be defined as its maximal subgraph in which each vertex has at least degree k . The k -core of a graph can be formed by recursively deleting all nodes with degree smaller than k . The cores of a graph form layers in which the $(k+1)$ -core is always a subgraph of the k -core. k -core decomposition can be used to analyse the cohesiveness of a network.

Network epidemics

Theoretical information diffusion models. Network epidemics: The majority of research efforts on modelling the flow of information and influence throughout networks has been done in the context of epidemiology, where the idea of epidemic is not limited to the spread of a virus but includes the spread of ideas, news, products or behaviour. The later are also described as *social contagion*.

The classical disease propagation models are based on the stages of a disease in a host: susceptible, infected and recovered. After recovery an actor can become immune (SIR) or once again susceptible (SIS). The initially infected nodes in our case would correspond to people who adopt a product without first receiving recommendations. In the SIR model, a susceptible person has a uniform probability β per unit time of becoming infected from any infected contact, while the infective individuals recover at some rate γ . The fractions s , i and r of individuals in the states S , I and R are determined by the differential equations:

$$\frac{ds}{dt} = -\beta is, \frac{di}{dt} = \beta is - \gamma i, \frac{dr}{dt} = \gamma i \quad (11)$$

In the case of SIS model the ‘cured’ individual goes back to the susceptible pool, thus the above differential equations are revised as follows:

$$\frac{ds}{dt} = -\beta is + \gamma i, \frac{di}{dt} = \beta is - \gamma i \quad (12)$$

The epidemic threshold

The *epidemic threshold* of the SIS and SIR models determines whether a disease can dominate or die out. In networks with power-law degree distribution there is no non-zero epidemic threshold so long as the exponent of the power-law is less than 3. Since most power-law networks satisfy this condition we expect diseases always to propagate in these networks, regardless of the transmission probability between individuals.¹⁴ However, there may be a non-zero threshold if the network is low-dimensional (rather than infinite) or if the network has high clustering coefficient.

Limitations of epidemiological models

Although these epidemiological models are useful in understanding the basic dynamics, their basic problem is that they assume that disease spreading depends only on a single parameter that specifies the infectiousness of the disease. This would mean that the entire population is equally susceptible to an idea or product purchase. Obviously, this assumption is unrealistic since diseases spread between individuals that have actual contact.

The Bass product diffusion model

Adoption of ideas or products: One of first product diffusion models was proposed by Bass.¹⁵ The Bass model predicts the number of people who will adopt an innovation over time. It is agnostic of the underlying network structure and it assumes that the adoption rate depends on the current proportion of population who have already adopted the innovation. The diffusion equation models the cumulative proportion of adopters in the population as a function of three parameters, the potential market, a co-efficient of innovation that expresses the intrinsic adoption rate, and a co-efficient of imitation p that determines the influence of social contagion. The cumulative adoption follows an S-curve, which means that initially the adoption is slow, increases exponentially and flattens at the end as shown in Figure 2.

The Bass model can provide a useful forecast at the aggregate level but it does not provide insights in the diffusion process.

The Linear Threshold Model

On the other hand, the Linear Threshold Model¹⁶ tries to predict the spread of products at the individual level. Each actor u in the network is assigned a threshold $t_u \in [0, 1]$ drawn from some probability distribution. Each actor u is influenced by each neighbour w according to a weight $b_{u,w}$, such that $\sum_w b_{u,w} \leq 1$. A node adopts a product if the sum of the connection weights of its neighbours who already adopted the product is greater than the threshold of the actor, namely

Figure 2: The Bass model of cumulative product adoption

$t \leq \sum_{adopters(i)} w_i$. After specifying the weights and the threshold for each actor, the diffusion process evolves deterministically in discrete steps.

A similar approach is followed by the Independent Cascade model.¹⁷ We start again with a set of initial adopters and the process unfolds in discrete steps. Whenever a neighbour v of an actor u adopts a product in a step t , it has single chance to also activate the actor u with probability $p_{u,v}$. If v succeeds, then w will become active in step $t+1$; but whether or not v succeeds, it cannot make any further attempts to activate u in subsequent steps.

The last two diffusion models are not independent from the network structure and their function depends on the selection of the initial adopters. These models give rise to the question of which individuals are more influential and better for the spread of a product.

The influence maximization problem

As described by the diffusion models, the decision that an individual takes of whether to adopt a product can be influenced, explicitly or implicitly, by their social contacts, as will be seen in the following section. In order to effectively employ 'word-of-mouth' recommendations, it is thus essential for companies to identify the so-called opinion leaders to target, assuming that influencing them will lead to a large cascade of further recommendations. More formally, the influence maximization problem is the following: given a probabilistic model for influence, can we determine a set individuals who will yield the largest expected cascade.

In many cases the most influential spreaders are not the best connected individuals

Early studies of scale-free networks identified the most connected people (hubs) as the key players in the large scale of spreading processes.¹⁴ Furthermore, in the context of social networks it is believed that the actors with the higher betweenness centrality exercise larger interpersonal influence.¹⁸ However, a recent study¹⁹ revealed that in many cases the most influential spreaders do not necessarily correspond to the best connected or more central individuals. Instead, by applying the SIR and SIS models in a series of online communities (livejournal.com, UCL CS email contact network, imdb.com), they find that a better metric for discovering the most influential spreaders is how close is the location of an individual to the core of the network. The core is defined by the k -shell decomposition (coreness). They also discover that when there are more than one initial spreaders, the distance between them is crucial to extending the cascade. Figure 3 shows how a hub produces a smaller cascade than an individual with fewer links but strategically placed close to the core of the network.

The strength of weak ties

An important contribution in understanding the spread of information is the theory of 'the strength of weak ties'.²⁰ This theory claims that individuals are often influenced by others with whom they have sparse or even random interactions. These interactions are labelled as 'weak ties' in contrast to the strong ties that an individual has with his close friends or family members. The significance of weak ties is based on the fact that they have the potential to bring together communities that otherwise are isolated from each other. Strong ties affect the spread of the information inside a closed circle of acquaintances, but weak ties facilitate the wider diffusion. This theory is also confirmed by

Figure 3: The extent of a cascade for different initial spreaders¹⁸

Goldenberg *et al.*¹⁷ where the authors — based on the Independent Cascade Model — show that weak ties are at least as influential as the strong ties. A direct indication is that in marketing systems based on referrals, the participants should be encouraged to refer people with whom they have a weak relationship, otherwise there is the risk that referrals will be limited to a small network subgroup.

Relevant experimental studies

Relevant research. The models presented in the previous section address the problem of how influence spreads in a network from a theoretical point of view and are based on assumed influence effects and not actual data. A number of experimental studies have been conducted to track the diffusion of information in online social networks. The difficulties involved in mining the Facebook data, however, mainly due to privacy issues, resulted in a relatively small set of studies directly tied to the platform, but with particularly important insights.

A critical question is whether Facebook follows the topological characteristics of other social networks that exhibit power-law degree distributions. Since it is not possible to obtain complete connectivity information, sampling techniques are used to obtain unbiased topological measurements.²¹ Sampling based on the Metropolis-Hasting random walk and the Re-weighted Random Walk reveal that node degree distribution exhibits a heavy tail (with high degree nodes occurring with higher probability) and not a power-law distribution as do most of the offline social networks. These results are also confirmed by Sala *et al.*²² which proposes the use of Pareto-Lognormal degree distribution for more accurate modelling of the Facebook network.

Facebook applications exhibit the property of preferential attachment

One of the first large-scale measurement studies on the Facebook OSN is presented in Nazir *et al.*²³ To obtain a rich dataset, the authors developed three Facebook applications whose popularity reached the top 1 per cent of Facebook applications with a combined user base of 8 million users. According to their results, the Facebook applications exhibit the *Preferential Attachment* property¹⁰ according to which the probability that a user will install an application is proportional to the current popularity of the application. This explains the fact that the application installations follow a power-law distribution. The preferential attachment derives directly from the news-feed of the

**A small number of
'power' users dominate
user interactions on
applications**

Facebook platform that informs a user for the activities of his friends. A second important observation is that the popularity of applications that cross a certain threshold of installations is not affected by sharp daily drops. In contrast, even small changes in the daily usage of less popular applications lead to significant changes in the application's rank. Also, the authors observe that for applications that cannot gain popularity during their initial deployment stage, it is very difficult to reach high ranking, indicating that this first stage is decisive for the future user base of an application. Finally, the authors observe that the user interactions also follow a power-law distribution. This result implies that a small number of 'power users' dominate the user interactions on applications, and are responsible for sustaining an application at a high rank.

Another important study on the diffusion of applications in the Facebook platform is presented in Onnela and Reed-Tsochas.²⁴ The authors study how social influence and collective behaviour affects the decision of individual users on whether to install an application. Their analysis is based on hourly data between June and August 2007 for 2,720 Facebook applications with a total of 104 million installations. Their results are surprising and show the existence of two distinct regimes of behaviour. Once applications cross a particular threshold of popularity, social influence becomes a highly determinant factor for the behaviour among the users, and leads some of the applications to extraordinary levels of popularity. Below this threshold, the collective effect of social influence appears to disappear almost entirely. These results demonstrate that social influence can spontaneously assume an on/off nature in an online social network, in a manner not observed in the offline world. Additional simulation results with synthetic time series show that this transition is not equivalent to a standard epidemic threshold.

**Facebook diffusion
is not the result of a
cascade**

Sun *et al.*²⁵ explore a dataset of 2,62,985 Facebook pages to provide an empirical investigation of diffusion and the effect of news feed broadcasting. This unique insight into diffusion on Facebook, built on data not normally accessible directly, offers similar conclusions to that previously stated. While long chains of up to 82 levels have been observed, Facebook diffusion is not or rarely the result of a large cascade. Instead, it exhibits diffusion patterns characterized by large-scale collisions of shorter chains: diffusion events are often related to publicly visible pieces of content that are introduced into a network from multiple sources, often merging into one large group of friends. Start nodes, at the source of the chains, constitute an average of 14.8 per cent of users for pages with over 1,000 fans. Using zero-inflated negative binomial regressions, the authors also find that maximum diffusion chains cannot be predicted based on demographics or number of friends.

Other approaches, not necessarily tied to Facebook, have aimed to uncover the structure of social network through the use of penetration data,²⁶ relying on dissemination patterns to estimate properties of the network (eg type of degree distribution such as gaussian/poisson,

The model correctly identifies network structures

uniform, scale-free or lognormal). Using this information as a basis, they define a growth model that more precisely estimates the magnitude of the contagion process. Through their experimental set-up, which relies on CD sales data, online movie data based on search query volume, and Friendster data, they conclude that underlying network structures can be correctly identified through the use of their model and observed ‘infection’ rates, such as search volumes over time.

Identified methodologies and limitations. There are a few key conclusions to take away from this initial set of related work. First content disseminated on Facebook appears to follow a similar pattern of diffusion to that of product diffusion: viral processes or word of mouth do not take place until a certain threshold of deployment has been reached. Only then can we observe the impact of social influence. This impact may be limited in its depth: for fan pages, it appears that a significant number of start nodes will have found the page independently, before some form of cascading takes place. Whether these patterns of use remain past the initial introduction of the content, or when the content appears to be no longer relevant, is not clear from these studies. Indeed, further work is needed to understand the complete content life cycle, from introduction to decline in popularity, in order to correctly determine the level of investment required throughout: this ‘viral life cycle’ is discussed in more detail in the section ‘Challenges and research opportunities’.

Active invitation has a different impact to passive broadcasting

While work has focused on independent subsets of the Facebook feature space, such as ‘fanning of pages’ or application installations, little has worked on comparing the impact of these features on the overall diffusion process. Actively allowing users to invite others to use an application has, for example, a different impact to passive broadcasting via the news feed, and this may significantly impact the marketing process in a way that is just as significant as the network structure. As such it is important to examine closely other approaches, which both look at the operation of specific Facebook features, such as the news feed population through ranking-based recommendations (section ‘Recommender systems’), or at the nature of the content itself (section ‘Semiotic and semantic analysis’).

Recommender systems

Recommendation can be integral to the social media experience

Overview. Recommender systems aim at identifying interesting items (eg books, movies, websites, conversations) for a given user based on their previously expressed interests and other data such as demographics and social links. Beyond the obvious use of recommendation systems to recommend items to purchase, their use can form an integral part of the social media experience: Facebook’s news feed algorithm EdgeRank, detailed in the section ‘Background: Facebook’, is an obvious example of a ranking-based approach to recommendation that aims to highlight conversations or other posts that

**Previous preferences
are likely to be similar
in the future**

a user is most likely to be interested in. It is for this reason that we examine in this section recommender systems, their operation and use in social media environments.

Collaborative filtering is one of the most popular techniques for this purpose,²⁷ relying on the basic idea that users who tend to have similar preferences in the past are likely to have similar preferences in the future. From there various approaches can be adopted: *model-based approaches* provide item recommendation by first developing a model of user ratings. The training samples are used to generate an abstraction, which is then used to predict ratings for items that are unknown to the user. Examples of probabilistic models that have been proposed in this regard include the work of Breese *et al.*,²⁸ who model item correlation using a Bayesian Network model, in which the conditional probabilities between items are maintained or²⁹ who compute the probability that users belong to particular personality types, as well as the probability that certain personality types like new items.

Memory-based approaches on the other hand are the most widely adopted. In these approaches, all user ratings are indexed and stored into memory. In the rating prediction phase, similar users or items are sorted based on the memorized ratings. Relying on the ratings of these similar users or items, a prediction of an item rating for a test user can be generated. Examples of memory-based collaborative filtering include user-based methods,²⁷ item-based methods³⁰ and combined methods.³¹

It is also worth mentioning content-based approaches, which take into account domain-specific knowledge during the process of generating prediction. Content-based³² recommender systems deal with domain specific knowledge in order to build on semi-structured information (eg the genre of the movie, the location of the user) and infer relationships between these structures, rather than limiting the analysis to the relationship between user and item.

**Users only check
the correctness of
recommendations at
the top of the list**

On the other hand, researchers also argued that understanding collaborative filtering as a rating prediction problem has some significant drawbacks. In many practical scenarios, such as the Amazon's book recommender system,³³ a better view of the task is of generating a top-N list of items that the user is most likely to like. The simple reason behind it is that users only check the correctness of the prediction at the top positions of the recommended list, they are not interested in items that would be predicted uninteresting to them, therefore low on the ranking list. Such ranking-based approaches have been explored in a wide range of works.^{34,35}

Recommender systems are tightly related to network-based marketing for the fundamental reason that both benefit from exploiting underlying affinities between users. As pointed out by Hill *et al.*,² however, few of these works take into account explicit links between users: users are instead linked based on shared purchases or similar ratings. Nevertheless, understanding and analysing affinities between users is key to building marketing messages that resonate with specific

users and improving the targeting process, which allows us to identify likely adopters.

We examine in the following subsection research that explores the use of recommendation systems in social media environments.

Gaining advantage by included social network links

Relevant research. The inclusion of social media links to improve collaborative filtering has been explored by a number of works. Zheng *et al.*³⁶ for example introduce social network links to complement traditional collaborative filtering approaches and predict rated items obtained from online communities, concluding that this provides a net advantage over simple by-item or by-user approaches. The method simply relies on weighted averages of ratings of the same items from self-selected friends. In addition they conclude that there is little evidence of peer influence.

Social network features are of greater value than other variables

Guo *et al.*³⁷ examine the inclusion and impact of social network features into the e-commerce world, alongside the standard recommendations, product reviews and search features that are now standard. They aim to determine whether social networks can shape purchasing decisions, the factors that influence the success of a recommendation, and the impact of social influence and reputation on the commercial activity. The data used as a basis for the research are obtained from a Chinese online marketplace called Taobao, which alongside the standard elements of functionality required to facilitate user to user sales, also includes an instant messaging platform. The analysis of trades, message exchanges and contact lists highlights a number of key findings. In particular, there exists a relationship between social proximity (measured by the number of mutual contacts), the frequency of message exchanges and the likelihood of trades. Using an approach based on directed triadic closure, they further refine their analysis to investigate the process of information passing, looking specifically at message exchanges between buyers prior to purchase from a seller. The overall analysis is used as a basis for a consumer choice prediction algorithm that uses a ranking-based machine learning approach to determine the overall likelihood that a buyer relies on one particular seller over another for a same given product with a 42 per cent likelihood of correctly picking the correct seller of ten potential options. The overall conclusion is that social network features are of greater value than other variables such as product price and seller rating in predicting purchase behaviour, and that buyer to buyer communication is one of the primary drivers of purchasing.

Important conclusions from this work relevant here are that social proximity, frequency of exchanges are all key indicators and have a potential impact on any marketing strategy. There are, however, key questions regarding the applicability to this on networks not directly related to e-commerce (eg Facebook) — and whether similar conclusions can be drawn on the virality of message diffusion.

Chen *et al.*,³⁸ define a combinational method of collaborative filtering, which aims to perform personalized community recommendations based on semantic and user information and a hybrid

Orkut

training strategy that combines Gibbs sampling and Expectation-Maximization. While we will discuss semantic approaches in more detail in the following section, the combined approach effectively relies on combining collections of users and words describing a community to identify content similarities. Using the social network Orkut as a source of community and user data, the authors collect information for over 3,00,000 users and 1,00,000 communities — though get limited insight into friend graphs due to privacy issues. The precision of the recommendations increase from 15 per cent to 27 per cent for users who have joined over 100 hundred communities using the hybrid approach and with further improvements as the number of communities joined increases.

Though the experimental approach relies on Orkut rather than Facebook, the result of these experiments highlights that where limitations may exist with regard to accessing social graphs, recommendation-based approaches and linguistic analysis approaches may serve to compensate for the lack of knowledge of the social graph. Communities and community descriptions serve as a mechanism for identifying potential topics of interest to users and further communities that could/should be targetted whether for the purpose of advertising or otherwise.

Twitter

Chen *et al.*³⁹ examine the use of ranking-based recommendation systems for recommending conversations in online social streams focusing particularly on Twitter, with obvious comparisons to EdgeRank. Their approach relies on thread length, topic relevance and tie strength as the main factors for exchanges. Thread length is included primarily on the basis that the longer the conversation length, and hence of richer content, the more likely a user is to be interested. Topic relevance relies on a bag of words approach, and a tf-idf (term frequency-inverse document frequency) weighting scheme (described in the following subsection) to identify topic interests in the form of frequently mentioned entities or terms. Finally the tie-strength is a characterization of the relationship between users, estimated by examining the frequency of past bi-directional communications between two users or their common friends. A survey of users showed that users expressed much more interest in algorithms including tie strength, as a measure of how efficient their approach was in identifying conversations and content relevant to them.

Collaborative filtering to predict the spread of information

Finally Zaman *et al.*⁴⁰ use collaborative filtering to predict the spread of information in social networks — looking specifically at Twitter data. Using data of who and what was shared in the form of a retweet (ie forwarding of a message posted by a user), the authors use a probabilistic collaborative filter model to predict future retweets at the micro-level. The features used as a basis for analysis were the name, number of followers, time frame and tweet content (main tweet words extracted). They obtained for this purpose 102 million retweets and a network of 50 million edges and 7.3 million distinct users and concluded that their model allows us to predict retweets up to a day in advance.

Models to predict items that will generate a response

Identified methodologies and limitations. While the ambition is not actually to build a recommender system, we can conclude that the approaches exploited for recommender systems are very likely to provide insights into the users — such as their interest networks. With a sufficient data set, models built on such systems are likely to help us identify not only items that are likely to appear in a user's news feed but also those that are most likely to generate some sort of response from that user. Alternatively, these data may help us identify potential 'likes' and user interests to complement missing demographics, which would serve to refine the marketing message. This is of particular worth as Facebook social plug-ins are increasingly deployed throughout the web.

Following Zaman's work,⁴⁰ it remains an open question whether such models can be used to determine the likelihood of sharing at a micro and macro-level in Facebook based on past interactions.

Semiotic and semantic analysis

Overview. As we have briefly touched upon in the previous section, while much work on contagion and recommendation systems examines the how of message propagation in networks, these do not examine the nature of the particular message, which facilitated its spread. As highlighted by Berger *et al.*:⁴¹

Focusing on network structure [...] and on the influence of social people provides little insight into why certain cultural items become viral while others do not.

To address this issue, various works have aimed to apply semiotic and semantic analysis to social networks in order to gain a clearer understanding of what may become viral, exploiting a wide range of techniques in the process, such as information retrieval and sentiment analysis.

Automatic indexing and retrieval

Information retrieval involves the automatic indexing and retrieval of information in some shape or form. Automated classification and analysis techniques are a must considering the volume of human generated data that can be obtained from social networks. Most of the popular search engines use techniques like term frequency weighting approaches to extract important terms out of documents and vector space approaches to rank the results according to their relevance.⁴² tf-idf weighting is a method to calculate the relevance of a term according to a document. The measure assumes that the importance of a term in a document corresponds to the number of appearances in this document as well in all documents of the whole corpus. A term is more important than another if the term has more occurrences in the document than another. The term frequency is combined with the inverse of the document frequency by multiplication. This allows common non-contextual words to be eliminated. Such information may be complemented with part-of-speech information using algorithms that associate terms with descriptive tags representing their relationship to

Opinion classification is usually two-way: positive and negative

other adjacent terms (eg verbs, nouns, adjectives, etc). In this way, a document, such as a conversation, wall post or other can be represented by or classified according to the identified main themes and topics that emerge.

Opinion mining and sentiment analysis build on similar concepts,⁴³ though are less concerned with the topic or factual content in it, but rather with the opinion expressed in a document. While issues such as subjectivity and viewpoints all come into play, opinion classification is usually framed as a two-way classification of positive and negative sentiment, which can then be applied at different levels: phrases, sentences, documents and collections of documents. Most sentiment analysis algorithms fall into two types, sentiment-lexicon-based algorithms and machine learning-based algorithms. Sentiment-lexicon-based algorithms construct functions based on features provided by sentiment lexicon such as term positive scores to calculate the polarity of the tested review. Machine learning-based algorithms typically include support vector machines:⁴⁴ these rely on training data labelled with sentiment values, which can take into account sentiment lexicon features to direct the support vector machines.⁴⁵ It should be stated that such approaches have been used to track a wider range of moods beyond the usual binary approach to positive and negative sentiment (eg calm, alert, sure, vital, kind and happy).⁴⁶

Travel images on Facebook

Beyond automated language analysis, several pieces of work have looked at pulling data in various forms (eg pictures) from social media networks to obtain insights into what draws response from a user community. White⁴⁷ examines for example travel images on Facebook surrounding tourism encounters to draw conclusions on which particular shots proved more popular in order to influence travel photography.

Though this may touch upon a wide range of fields (socio linguistics, psychology, etc), we focus here on that most relevant to social networks.

Relevant research. Greetham *et al.*⁴⁸ examine the spread of positive and negative effect in social networks, in an experiment involving 100 university students logging all interactions with peers and their general positive and negative affect. An analysis of network dynamics based on a stochastic actor-based model concludes that a relationship exists between better health, greater negative affect and a greater number of interactions.

New York Times

Similarly, Berger *et al.*⁴¹ examine generally the type of content that is shared in social networks and the psychological processes that act as a selection mechanism and shape the process of virality. For this purpose, they examine the sharing of 7,000 *New York Times* articles by email, relying on automated sentiment analysis and human coders to identify more complex feelings associated with particular articles (utility, anger, anxiety, etc). Beyond content characteristics, they also control for external factors such as the appearance of the article on homepage, release timing, writing complexity and article length. The

summary of results indicates that surprising, practically useful and evoking emotional response were more likely to be emailed. They also, however, conclude that articles that spent more time on the homepage were more likely to be shared, as well as those written by more famous authors. In short, we may conclude that both affect and the mechanisms behind social transmission (the specific sharing mechanism and the features of the product being shared) have an impact on the virality of an item.

Comparing buzz

Doshi *et al.*⁴⁹ use a combination of social network analysis and sentiment analysis to attempt to predict trends, focusing particularly on movie prices. By comparing buzz, represented by blog betweenness and other social network analysis metrics, sentiment metrics obtained from discussions in forums, and box office performance data, they use multilinear and non-linear regression to attempt to predict final box office return. Though the social networks are constructed from semantic networks, using links and references between bloggers, over the standard mutual links established in most social media platforms, the combined use of social network analysis and semantic analysis, provides an interesting basis for predicting content that will prove popular. Sentiment analysis relies on a *bag of word* approach, though dynamically adapted to different movie genres. Such an approach results in a correct prediction of movie success or flop 80 per cent of the time. The research provides an interesting insight into the popularity of online content. The overall strategy allows to infer a potential link between content and diffusion in social networks.

Bag of word approach

Public mood and the Dow Jones Industrial Average

Alternatively, Bollen *et al.*⁴⁶ rely on Twitter data to determine whether mood expressed correlates with stock market activity. Using mood tracking tools that measure moods across dimensions such as calm, alert, sure, vital, kind and happy, based on expressions in the twitter message content (eg I am feeling) and a large data set (9,853,498 tweets posted by approximately 2.7 million users), they establish a relationship between public mood states and the Dow Jones Industrial Average. Using Granger causality analysis and a Self-organizing Fuzzy Neural Network, they find an accuracy of 87.6 per cent in predicting the daily up and down changes in the closing values of the Dow Jones Industrial Average.

A similar approach has been used to forecast audience increase on Youtube.⁵⁰ On the basis of the notion that quantifying the standing and perception of a user can be achieved by measuring his/her in-degree on a given social web platform, Rowe posits that a relationship exists between the behaviour of the community and subscribers count. Using a multiple linear regression approach, and a data set involving 2,000 uploaded videos monitored on a regular basis, the author examines whether current user in-degree (subscriptions to the user), out-degree (channels followed by the user), view count, post count, favourite count have a bearing on the increase in subscribers. Their results indicate that the post view counts is found to have a significant correlation with in-degree, as well as the number of times the content is favoured. An observation

is made, however, that there appears to be a negative correlation between increases in participation in the community by the user (more content uploaded, more video views), implying that excessive participation may have an effect on reputation. The authors believe that additional conclusions may be drawn from this work through the linguistic analysis of comments/titles.

**Crucial to understand
the type of content
likely to be shared**

Identified methodologies and limitations. As organizations aim to use social media to increase customer engagement, the effectiveness of these approaches hinges on creating content that people want to share. While some have argued that there may be no formula for what becomes viral,⁵¹ the work discussed here demonstrates that both the nature of the content and the process by which the content is introduced (eg position of links on a website) may have a high impact on the propagation process. It is hence crucial to understand the type of content that is likely to be shared by consumers, and whether we can infer from the initial response obtained whether this content is likely to further propagate or reach new customers. Fan pages and wall posts in Facebook provide a wealth of textual content that can be analysed, both for sentiment, or as described previously for the purpose of recommendations and user interest analysis. Exploiting such findings may serve to construct user engagement processes, which are much more effective than current approaches.

**Little insight into
Facebook sharing**

Summary

In this section, we have explored a wide range of methodologies and approaches that allow us to measure the impact of structure of the Facebook social graphs on the diffusion of messages and identify key influencers. We have also explored techniques that aim to provide key insights into the demographic and interest profiles of users, exploring the use of recommendation systems and linguistic analysis in social media environments. Though some level of work has been exploring the reasons why particular items and the specific characteristics exhibited by particular types of messages are more likely to propagate, we find that there is still little insight on this topic — particularly with respect to the specific nature of the Facebook features that may lead to messages being shared via the news feed or otherwise.

Monitoring Facebook metrics, described in the section ‘Background: Facebook’, allows us to examine the effects of passive broadcasting of actions via the news feed as well as the effect of the active involvement of users and empirically evaluate the impact of continuous customer engagement beyond initial customer acquisition. The observations of repeat interactions with an application or fan page may result in triggering interest, or dislike, in a way that might have further effect on the propagation of content in the networked environment. In this respect, we may begin to model the complete life cycle of viral content on Facebook, as will be discussed in the following section.

Challenges and research opportunities

Viral life cycle

We can summarize the challenges surrounding network-based marketing as follows:

- *Diffusion measurement and forecasting*: Can we reliably estimate the spread and diffusion of content on specific marketing platforms?
- *Impact measurement and forecasting*: Can we reliably estimate the adoption of a marketing message?

The diffusion model

Work discussed in the previous section has begun to provide us with an answer to these questions. It also allows us to draw strong parallels between the diffusion of messages in a social network environment and the product diffusion model⁵² developed by Frank Bass, and discussed in the section ‘Related work and state of the art’. The diffusion model and product life cycle, illustrated in Figure 4, have been widely influential in marketing and management science, and describe the progressive phases traversed by a product, from introduction to growth, maturity and decline, generally impacting the marketing strategy and the marketing mix. This may be associated with the Roger’s adoption curve,⁵³ which describes different product adoption groups, from innovators and early adopters, early majority to late majority and laggards.

The viral life cycle

Though as previously this model is intended as a generalization, obvious links can be established with the ‘viral life cycle’, based on our examination of related work. An initial product introduction period may require active promotion and reaching out to influential users, or users situated at the centre of desirable networks, either through advertisement, or influencer outreach. Alternatively the introduction

Figure 4: Product diffusion and adoption life cycle

The maturity phase

may be the result of events external to the platform, such as current events and publication in other popular forms of media.

The growth phase is that reached when the viral process begins to amplify the diffusion process, either through messaging, news feed or other social platform-specific features, reaching in the process an early majority. From there, a maturity phase begins, where the strong growth diminishes, but the content remains popular, and has become commonplace. Finally there is a decline phase, where effectively interest in the content begins to diminish, either because it is no longer relevant, lost novelty value or no longer benefits from marketing and community building efforts.

Our review of related works shows that there are a wide range of factors that have a potential impact on both the diffusion and adoption process: user connections, influence, behaviour and position in the network all have a potential impact on how a message propagates in a network. The nature of the content itself is extremely important in determining the likelihood that one may adopt or pass on the message.

Features of the social network can ensure growth

However, while much of the work reviewed examines the initial introduction and growth stage of content diffusion in social networks, little has looked at the stages beyond, or specifically at how the features of the social network can allow one to ensure that growth is retained or the general decline postponed.

A clear understanding of the complete viral life cycle can effectively allow us to determine

- When active investment and promotion is required to bring attention to the content in order for word of mouth to take place and enter the growth phase.
- The user engagement process and active community building required to ensure that we reach the early and late adopters and generally elongate the maturity phase.
- When continuous investment is no longer sensible as there has been a failure to progress to the natural growth phase, or the decline phase has begun.

Facebook features

An analysis of Facebook metrics described in the section 'Facebook graph objects' can provide us with the ability to determine the current stage of the life cycle process and identify the most effective approaches used to retain user interest beyond the initial introduction and growth phase. The Facebook feature space, however, remains relatively large and it is important to decompose it in order to identify those key metrics most relevant to the specific dimension we aim to examine.

Monitoring domains

Revisiting the section 'Facebook graph objects', we distinguish fan pages, applications built on the Facebook platform and general content, due to the strong differences that exist in how we approach both the

Table 2: Top 10 Facebook Brand pages in the food and beverage industry

Brand	Fans	Posts	Comments	Likes	Total interactions	Contributing users	Core
Coca-Cola	26,414,466	7,846	4,007	136,196	148,049	8,582	1477
Starbucks	21,701,785	8,371	6,209	221,393	235,973	10,576	2008
Skittles	16,339,447	1,947	1,670	181,416	185,033	2,917	398
Pringles	12,825,199	911	472	20,572	21,955	1,201	108
Ferrero Rocher	10,298,685	8,083	1,345	48,437	57,865	8,358	114
Dr Pepper	8,874,222	4,672	7,143	317,380	329,195	8,765	1230
Nutella	8,220,804	2	99	14,864	14,965	93	8
McDonald's	8,140,161	0	0	0	0	0	0
Taco Bell	6,426,008	3,316	5,177	108,471	116,964	5,555	1231
5 Gum	4,619,013	1,345	1,488	4,828	7,661	1,942	422

monitoring process, and the expected impact of network-based marketing.

Fan pages

The real-world impact of liking

Fan pages provide a wealth of information that is publicly available, which can be complemented with relevant administrative data. The real-world impact of 'liking' a fan page and participating in brand-related conversations may often be difficult to quantify; however, some studies such as Syncapse⁵⁴ have attempted to associate a monetary value to fans. In this cited study, which relies on questionnaires collected from 4,000 users in North America in 2010, the findings highlight that fans spend an additional \$71.84 on products for which they are fans compared to those who are not fans. They are 28 per cent more likely than non-fans to continue using the brand, and are 41 per cent more likely than non-fans to recommend a fanned product to their friends. The survey does highlight, however, that fan value can vary widely from page to page and fan to fan.

The process of 'fanning'

More direct means of tracking return on investment may be exploited: the dissemination of links on the page, referral codes, etc can provide a direct path between fan acquisition and purchase. However, the creation and promotion of a fan page often constitutes a branding exercise, and as such the process of 'fanning' a brand often constitutes an adoption in and of itself.

While active acquisition of fans may take place via traditional advertisements, cross platform promotion etc, the primary mode of diffusion of fan page activity within Facebook takes place via passive news feed broadcasting. In this context, numerous metrics can be taken into account to actively measure diffusion and adoption, as described in the section 'Facebook graph objects'.

Highly active users are passionate about the brand

We highlight in Table 2 some key metrics calculated from a selected number of pages. The table provides key metrics for fan pages related to the food and beverage industry over a period of a month from 6 April to 6 May 2011. Here, interactions are defined as any form of action by users, such as posting, commenting or liking a page. These data give an indication of how actively users were interacting with the

page, causing posts about the brand to appear in their news feeds. Contributing users are defined as members conversing on the brand page. Core active users are defined as users who post and create content on a brand page more frequently than the average user (relative to the page). This highly active group of users are passionate about the brand.

These data constitute a basic set of metrics that can be then complemented with linguistic analysis (what is the main sentiment/themes of the page) or administrative information (what are the demographics/impressions/etc) in order to answer the following questions:

- *Can we determine the current stage of the fan page with respect to the 'viral life cycle'?* Is further promotion required? Is the page in a continuous period of growth? Is it possible to acquire new fans or have we reached a stage of maturity and decline?
- *Can we maximize the impact of passive broadcasting through active user engagement?* Are communities with a large active core with regular repeating interactions more likely to grow? Should specific subsets of users be targeted based on their specific characteristics (links, position in the network cluster, etc)?
- *Can fan page content affect fan acquisition?* Can strong positive sentiment lead to better fan acquisition? May post quality or emerging themes affect the likelihood that a user becomes a fan?
- *What methodologies can we exploit to facilitate active user engagement?* Does posting at particular times of the day help engagement? Is regular posting by the page owner more likely to result in further engagement or may it hinder participation? Can we identify key themes or topics that resonate with the user?

Buzz may lead to increases in activity

Beyond this, we may wish to examine what external factors will have a strong impact on fan acquisition. General buzz or mentions of topics/fan pages on other platforms or the general web may lead to increases in activity, and monitoring activity external to the page may be of great value. We may also wish to compare specific features of Facebook in acquiring fans, such as the use of Facebook advertisements compared to passive news feed broadcasting.

The overall conclusions of such work would be to identify those metrics that are most likely to affect page propagation and fan acquisition and define an analytical model for the diffusion of fan pages that would allow us to forecast page adoptions based on the various properties exhibited by the page.

Establishing correlations between such metrics and fan page growth would most likely require data from numerous fan pages at various stages of the overall life cycle. Administrative access would be of significant value in this instance, providing us with news feed impressions and similar administrative access only metrics. Referring to Sun *et al.*²⁵ as a guide, reliably establishing such correlations may require a data set of up to several thousand pages. We may, however,

be content on testing user engagement strategies on multiple large pages, whose fan count has surpassed the previously identified threshold of adoption. As Facebook history can be available up to multiple years, regular monitoring may not be required to assist in this task; however, the ability to harvest posts and conversation threads for large number of pages would be.

Facebook applications

Exploiting a wider range of features

We discuss here the potential monitoring of Facebook specific widgets, games or other applications that are built on the social media platform. Facebook-based applications provide us with similar metrics to that of Fan pages, but with a wider range of potential features to be exploited and a potentially more direct line between application installs and return on investment. The application may be used to sell a wide variety of items, advertise or use micropayments in the form of Facebook credits or other⁵⁵ in order to purchase premium features or services. Through these means, application developers Zynga is, for example, expected to earn an estimated \$1.8 bn⁵⁶ in revenue in 2011.

Application developers have the opportunity to request permission to access various aspects of the user profile, including friend network, likes or other, providing detailed insights into the user behaviour and status. This may also serve to identify a clear propagation path, detailing the links between potential users of the application.

In this context, the process of adoption would constitute an application installation, and the primary concern of a developer may be to maximize the number of installations. A further step may go beyond and examine the specific features of the application itself, their use and overall ROI obtained. In this instance, we are concerned primarily with the specific features of the application itself that will facilitate propagation.

Similar questions to the propagation of fan pages emerge:

- *Can we determine the current stage of the application with respect to the 'viral life cycle'?*
- *Can we maximize the impact of passive broadcasting by ensuring regular application use?*
- *Which application features are more likely to increase propagation?*

Customised applications

The use of customized applications, as explored in Aral and Walker,⁷ provides an opportunity to test different application features on different user groups. Beyond comparing application invites to passive broadcasting, there exists an opportunity to examine and compare application features that specifically target other users. Another potential source of investigation is the reduction of impact caused by repeated updates — leading to potential application uninstalls or deletion of updates.

The overall conclusion of such work would be to define a model for diffusion of application installations, which would allow us to actively predict potential installations and determine patterns of use.

In this instance, monitoring multiple applications at different stages of the application life cycle may prove of great value. In particular, the side by side evaluation of specific features may be of greater value when examining widely deployed applications with multiple feature sets. Onnela and Reed-Tsochas²⁴ examine over 2,000 applications, though largely deployed applications may be of great value. Examining potential network structures, due to the potentially large data sets, may only be feasible on relatively small data sets, with Sun *et al.*²⁵ for example limiting their examinations to ten pages.

General content

We refer here in this section to any potential piece of content that is potentially shared between users, either through wall to wall messaging, link sharing or other direct interaction feature. This may be a link, brand or product mention discussed in any open setting (public wall, public exchange, etc).

Pages may express dissatisfaction with a brand

Monitoring of mentions may take place in many forms, through active searching for open status and monitoring of fan pages. This approach also may allow the identification of pages created by users independently from a brand. These may, for example, serve to express dissatisfaction surrounding a brand product, such as the user initiated Facebook group 'Bid to get a Fairtrade Mars Bar'. As a whole, both public status updates and monitoring of related pages may serve to identify increases in activity surrounding a brand or a product, and identify potential propagation paths. However, as few data are available regarding the percentage of Facebook profiles that may be open, it may be difficult to estimate how representative searches for mentions of brands or topics may be with respect to the overall number of conversations taking place on Facebook as a whole.

Referral links may identify individual users

An alternative approach is to rely on links to third-party websites. Detailed site analytics can provide source information, traffic analysis and other information relevant to propagation. The additional inclusion of referral links may serve to identify individual users and subgroups of users and the potential propagation path. Generally the use of penetration data can prove effective in network reconstruction, as discussed in Dover *et al.*²⁶

In this context it becomes possible to explore the particular impact of specific Facebook features: the active sharing of links for example on Facebook may have a different impact on propagation than the use of social plug-ins, where external sites can embed 'like' buttons and the like on their page, or alternatively the use of Facebook advertisements.

Conclusion

In this document, we have examined existing work in network-based marketing in social media, with particular focus on Facebook. Conclusions from this work indicate that network structure, themes and user profiles all have a significant impact on the diffusion and

Little examination of viral content's

adoption of marketing messages in social media, and as such various methodologies ranging from linguistic analysis to recommender systems can be exploited to explore and anticipate diffusion.

We have also detailed in the section 'Challenges and research opportunities' potential venues for further exploration, which are deemed particularly important in understanding how best to exploit a social media platform such as Facebook for marketing purposes. In particular we have found that there is little work examining the overall life cycle of viral content, with much work limited to initial growth. Understanding the different stages of the life cycle, specifically introduction, growth, maturity and decline, is important in helping us understand when we may no longer benefit from the viral process or whether additional investment is required to delay a potential decline.

Do brands benefit from user engagement?

Much of the existing bulk of work is concerned with identifying network structure and influences, but little examines the particular features of Facebook, such as the news feed, which are more likely to result in the propagation of content. Significant investment has gone into creating and maintaining Facebook fan pages and applications from a wide range of brands; however, the actual impact of different communication approaches with the end user has not been evaluated. In short, the question as to whether or not brands benefit from active and continuous user engagement and the building of communities on the Facebook platform has not been explored. Little insight has been provided on what are the most effective approaches to user engagement, and whether themes and content have a potential impact on customer acquisition.

We have detailed in this context three potential venues of further exploration and examined the monitoring requirements of these domains: fan pages, application and general content. Fan pages, due to the wealth of public information, may prove to be most valuable in beginning to model propagation on the Facebook platform.

References

1. eMarketer. (2010) 'Social network ad spending to approach \$1.7 billion this year' available at <http://www.emarketer.com/Article.aspx?R=1007869>, accessed 1 June 2012.
2. Hill, S., Provost, F. and Volinsky, C. (2006) 'Network-based marketing: Identifying likely adopters via consumer networks', *Statistical Science*, Vol. 21, No. 2, pp. 256–276.
3. Facebook. (2012) 'Key facts', available at <http://newsroom.fb.com/content/default.aspx?NewsAreaId=22>, accessed 1 June 2012.
4. O'Neill, N. (2009) 'AllFacebook.com. Zynga rocketing past 200 million users', available at http://allFacebook.com/zynga-rocketing-past-200-million-users_b8734, accessed 1 June 2012.
5. Tsotsis, A. (2010) 'TechCrunch. Facebook, by the numbers', available at <http://techcrunch.com/2010/08/17/Facebook-graphic/>, accessed 1 June 2012.
6. Kincaid, J. (2010) 'TechCrunch. The secret sauce that makes Facebook's news feed tick' available at <http://techcrunch.com/2010/04/22/Facebook-edgerank/>, accessed 1 June 2012.
7. Aral, S. and Walker, D. (2010) 'Creating social contagion through viral product design: A randomized trial of peer influence in networks', *Management Science*, Vol. 57, No. 9, pp. 1623–1639.
8. Costa, L., Rodrigues, F., Travieso, G. and Villas Boas, P.R. (2007) 'Characterization of complex networks: A survey of measurements', *Advances in Physics*, Vol. 56, No. 1, pp. 167–242.

9. Erdős, P. and Rényi, A. (1959) 'On random graphs I', *Publicationes Mathematicae Debrecen*, Vol. 6, p. 290.
10. Barabási, A. and Albert, R. (1999) 'Emergence of scaling in random networks', *Science*, Vol. 286, No. 5439, pp. 509–512.
11. Newman, M.E.J. (2003) 'The structure and function of complex networks', *SIAM Review*, Vol. 45, No. 2, pp. 167–256.
12. Bollobás, B. and Riordan, O. (2004) 'The diameter of a scale-free random graph', *Combinatorica*, Vol. 24, No. 1, pp. 5–34.
13. Freeman, L. (1977) 'A set of measures of centrality based on betweenness', *Sociometry*, Vol. 40, No. 1, pp. 35–41.
14. Pastor-Satorras, R. and Vespignani, A. (2001) 'Epidemic spreading in scale-free networks', *Physical Review Letters*, Vol. 86, No. 14, pp. 3200–3203.
15. Bass, F.M. (2004) 'A new product growth for model consumer durables', *Management Science*, Vol. 50, No. 12, p. 1825.
16. Granovetter, M. (1978) 'Threshold models of collective behavior', *The American Journal of Sociology*, Vol. 83, No. 6, pp. 1420–1443.
17. Goldenberg, J., Libai, B. and Muller, E. (2001) 'Talk of the network: A complex systems look at the underlying process of word-of-mouth', *Marketing Letters*, Vol. 12, No. 3, pp. 211–223.
18. Freeman, L.C. (1979) 'Centrality in social networks conceptual clarification', *Social networks*, Vol. 1, No. 3, pp. 215–239.
19. Kitsak, M., Gallos, L.K., Havlin, S., Liljeros, F., Muchnik, L., Stanley, H.E. and Makse, H.A. (2010) 'Identification of influential spreaders in complex networks', *Nature Physics*, Vol. 6, No. 11, pp. 888–893.
20. Friedkin, N. (1980) 'A test of structural features of Granovetter's strength of weak ties theory', *Social Networks*, Vol. 2, No. 4, pp. 411–422.
21. Gjoka, M., Kurant, M., Butts, C.T. and Markopoulou, A. (2010) 'Walking in Facebook: A case study of unbiased sampling of OSNS', in R. Boutaba and M. Chatterjee (eds.) *INFOCOM, 2010 Proceedings IEEE*, IEEE, San Diego, California, pp. 1–9.
22. Sala, A., Zheng, H., Zhao, B.Y., Gaito, S. and Rossi, G.P. (2010) 'Revisiting the powerlaw degree distribution for social graph analysis', in A. Richa and R. Guerraoui (eds.) *Proceeding of the 29th ACM SIGACT-SIGOPS Symposium on Principles of Distributed Computing*, ACM, Zurich, Switzerland, pp. 400–401.
23. Nazir, A., Raza, S. and Chuah, C.N. (2008) 'Unveiling Facebook: A measurement study of social network based applications', in K. Papagiannaki and Z.L. Zhang (eds.) *Proceedings of the 8th ACM SIGCOMM Conference on Internet Measurement*, ACM, Vouliagmeni, Greece, pp. 43–56.
24. Onnela, J.P. and Reed-Tsochas, F. (2010) 'Spontaneous emergence of social influence in online systems', *Proceedings of the National Academy of Sciences*, Vol. 107, No. 43, pp. 18375.
25. Sun, E.R., Marlow, I.C. and Lento, T. (2009) 'Gesundheit! Modeling contagion through Facebook news feedin Proc. of International AAAI Conference on Weblogs and Social Media.
26. Dover, Y., Goldenberg, J. and Shapira, D. (2009) 'Uncovering social network structures through penetration data(working paper)', available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1526484, accessed 1 June 2012.
27. Herlocker, J.L., Konstan, J.A., Borchers, A. and Riedl, J. (1999) 'An algorithmic framework for performing collaborative filtering', in F. Gey, M. Hearst and R. Tong (eds.) *Proceedings of the 22nd annual international ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM, Berkeley, California, pp. 230–237.
28. Breese, J.S., Heckerman, D. and Kadie, C. (1998) 'Empirical analysis of predictive algorithms for collaborative filtering', in G.F. Cooper and S. Moral (eds.) *Proceedings of the 14th conference on Uncertainty in Artificial Intelligence*, Morgan Kaufmann, Madison, pp. 43–52.
29. Pennock, D.M., Horvitz, E., Lawrence, S. and Giles, C.L. (2000) 'Collaborative filtering by personality diagnosis: A hybrid memory-and model-based approach', in C. Boutilier and M. Goldszmidt (eds.) *Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence*, Citeseer, San Francisco, California, pp. 473–480.

30. Deshpande, M. and Karypis, G. (2004) 'Item-based top-n recommendation algorithms', *ACM Transactions on Information Systems (TOIS)*, Vol. 22, No. 1, pp. 143–177.
31. Wang, J., De Vries, A.P. and Reinders, M.J.T. (2006) 'Unifying user-based and item-based collaborative filtering approaches by similarity fusion', in S. Dumais, D. Hawking and K. Järvelin (eds.) *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM, Seattle, WA, pp. 501–508.
32. Pazzani, M. and Billsus, D. (2007) 'Content-based recommendation systems', in Brusilovsky, P., Kobsa, A. and Nejdl, W (eds) *The Adaptive Web*, Springer, Germany, pp. 325–341.
33. Linden, G., Smith, B. and York, J. (2003) 'Amazon.com recommendations: Item-to-item collaborative filtering', *Internet Computing, IEEE*, Vol. 7, No. 1, pp. 76–80.
34. Guan, Z., Bu, J., Mei, Q., Chen, C. and Wang, C. (2009) 'Personalized tag recommendation using graph-based ranking on multi-type interrelated objects', in M. Sanderson, C.X. Zhai and J. Zobel (eds.) *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM, Boston, MA, pp. 540–547.
35. Coyle, L. and Cunningham, P. (2004) 'Improving recommendation ranking by learning personal feature weights', in Roth-Berghofer, T., Göker, M.H., Güvenir, H. (eds) *Advances in Case-based Reasoning: 8th European Conference, ECCBR 2006, Fethiye, Turkey, September 4–7, 2006, Proceedings (Lecture Notes in Computer Science/Lecture Notes in Artificial Intelligence)*, Springer, Germany, pp. 7–29.
36. Zheng, R., Wilkinson, D. and Provost, F. (2012) 'Social Network Collaborative Filtering (October 2008)Stern, IOMS Department, CeDER', available at SSRN: <http://ssrn.com/abstract=1303924>, accessed 1 June 2012.
37. Guo, S., Wang, M. and Leskovec, J. (2011) 'The role of social networks in online shopping: Information passing, price of trust, and consumer choiceArxiv preprint arXiv:1104.0942', available at <http://arxiv.org/abs/1104.0942v1>, accessed 1 June 2012.
38. Chen, W.Y., Zhang, D. and Chang, E.Y. (2008) 'Combinational collaborative filtering for personalized community recommendation', in B. Liu and S. Sarawagi (eds.) *Proceeding of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, Las Vegas, NV, pp. 115–123.
39. Chen, J., Nairn, R. and Chi, E. (2011) 'Speak little and well: Recommending conversations in online social streams', in D. Tan (ed.) *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems*, ACM, Vancouver, BC, pp. 217–226.
40. Zaman, T.R., Herbrich, R., Van Gael, J. and Stern, D. (2010) 'Predicting information spreading in twitter', in H. Wallach and J.W. Vaughan (eds.) *Workshop on Computational Social Science and the Wisdom of Crowds*, NIPS, Whistler, Canada.
41. Berger, J. and Milkman, K.L. (2009) 'Social transmission and viral culture(working paper)', available at http://faculty.insead.edu/marketing_seminars/Seminars%202009-10/Marketing%20Camp%202010/Jonah%20Berger/Berger_Virality_paper%201.pdf, accessed 1 June 2012.
42. Manning, C.D., Raghavan, P. and Schütze, H. (2008) *Introduction to Information Retrieval*, Cambridge University Press, Cambridge.
43. Pang, B. and Lee, L. (2008) 'Opinion mining and sentiment analysis', *Foundations and Trends in Information Retrieval*, Vol. 2, No. 1–2, pp. 1–135.
44. Joachims, T. (1999) 'Making large-scale support vector machine learning practical', in Schölkopf, B., Burges, C., Smola, A. (eds) *Advances in Kernel Methods: Support Vector Learning*, MIT Press, USA, pp. 169–184.
45. Esuli, A. and Sebastiani, F. (2005) 'Determining the semantic orientation of terms through gloss classification', in O. Herzog (ed.) *Proceedings of the 14th ACM International Conference on Information and Knowledge Management*, ACM, Bremen, Germany, pp. 617–624.
46. Bollen, J., Mao, H. and Zeng, X. (2011) 'Twitter mood predicts the stock market', *Journal of Computational Science*, Vol. 2, No. 2011, pp. 1–8.
47. White, L. (2009) 'Facebook, friends and photos: A snapshot into social networking', in *Tourism Informatics: Visual Travel Recommender Systems, Social Communities, and User Interface Design*, Information Science Reference, USA, p. 115.
48. Greetham, D.V., Hurling, R., Osborne, G. and Linley, A. (2010) 'Social networks and positive and negative affectCurrently under submission to the 7th conference on Applications

- of social network analysis', available at http://www.asna.ch/fileadmin/user_upload/2010/Papers/Vukadinovic-et-al_paper.pdfPreprint, accessed 1 June 2012.
49. Doshi, L., Krauss, J., Nann, S. and Gloor, P. (2010) 'Predicting movie prices through dynamic social network analysis', *Procedia — Social and Behavioral Sciences*, Vol. 2, No. 4, pp. 6423–6433.
 50. Rowe, M. (2011) 'Forecasting audience increase on YouTube', in G. Antoniou, M. Grobelnik, E. Simperl, B. Parsia, D. Plexousakis, P. De Leenheer and J. Pan (eds.) *Workshop on User Profile Data on the Social Semantic Web, 8th Extended Semantic Web Conference 2011 (ESWC 2011)*, 30 May 2011, Heraklion, Greece.
 51. Salganik, M.J., Dodds, P.S. and Watts, D.J. (2006) 'Experimental study of inequality and unpredictability in an artificial cultural market', *Science*, Vol. 311, No. 5762, p. 854.
 52. Frank, M.B. (1969) 'A new product growth for model consumer durables', *Management Science*, Vol. 15, No. 5, pp. 215–227.
 53. Rogers, E.M. (1995) *Diffusion of Innovations*, (4th edition) The Free Press, USA.
 54. Syncapse. (2010) 'The value of a Facebook fan', available at <http://syncapse.com/value-of-a-facebook-fan-an-empirical-review/#more-2200>, accessed 1 June 2012.
 55. Facebook. (2012) 'Facebook credits', available at <http://www.Facebook.com/credits>, accessed 1 June 2012.
 56. Kosman, J. and Sloane, G., New York Post (2011) 'Look at Zynga's Book\$\$', available at http://www.nypost.com/p/news/business/look_at_zynga_book_mIo4hDhRprH3MfuNXUcnDM, accessed 1 June 2012.