Differences in the Mechanics of Information Diffusion Across Topics: Idioms, Political Hashtags, and Complex Contagion on Twitter

Daniel M. Romero
Cornell University
Ithaca, NY
dmr239@cornell.edu

Brendan Meeder
Carnegie Mellon University
Pittsburgh, PA
bmeeder@cs.cmu.edu

Jon Kleinberg
Cornell University
Ithaca, NY
kleinber@cs.cornell.edu

ABSTRACT

There is a widespread intuitive sense that different kinds of information spread differently on-line, but it has been difficult to evaluate this question quantitatively since it requires a setting where many different kinds of information spread in a shared environment. Here we study this issue on Twitter, analyzing the ways in which tokens known as hashtags spread on a network defined by the interactions among Twitter users. We find significant variation in the ways that widely-used hashtags on different topics spread.

Our results show that this variation is not attributable simply to differences in "stickiness," the probability of adoption based on one or more exposures, but also to a quantity that could be viewed as a kind of "persistence" — the relative extent to which repeated exposures to a hashtag continue to have significant marginal effects. We find that hashtags on politically controversial topics are particularly persistent, with repeated exposures continuing to have unusually large marginal effects on adoption; this provides, to our knowledge, the first large-scale validation of the "complex contagion" principle from sociology, which posits that repeated exposures to an idea are particularly crucial when the idea is in some way controversial or contentious. Among other findings, we discover that hashtags representing the natural analogues of Twitter idioms and neologisms are particularly non-persistent, with the effect of multiple exposures decaying rapidly relative to the first exposure.

We also study the subgraph structure of the initial adopters for different widely-adopted hashtags, again finding structural differences across topics. We develop simulation-based and generative models to analyze how the adoption dynamics interact with the network structure of the early adopters on which a hashtag spreads.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous

General Terms

Theory, Measurement

Keywords

Social media, social contagion, information diffusion

1. INTRODUCTION

A growing line of recent research has studied the spread of information on-line, investigating the tendency for people to engage in activities such as forwarding messages, linking to articles, joining groups, purchasing products, or becoming fans of pages after some number of their friends have done so [1, 4, 7, 9, 15, 20, 22, 23, 29]. The work in this area has thus far focused primarily on identifying properties that generalize across different domains and different types of information, leading to principles that characterize the process of on-line information diffusion and drawing connections with sociological work on the *diffusion of innovations* [27, 28].

As we begin to understand what is common across different forms of on-line information diffusion, however, it becomes increasingly important to ask about the sources of variation as well. The variations in how different ideas spread is a subject that has attracted the public imagination in recent years, including best-selling books seeking to elucidate the ingredients that make an idea "sticky," facilitating its spread from one person to another [11, 16]. But despite the fascination with these questions, we do not have a good quantitative picture of how this variation operates at a large scale.

Here are some basic open questions concerning variation in the spread of on-line information. First, the intuitive notion of "stickiness" can be modeled in an idealized form as a probability — the probability that a piece of information will pass from a person who knows or mentions it to another person who is exposed to it. Are simple differences in the value of this probability indeed the main source of variation in how information spreads? Or are there more fundamental differences in the mechanics of how different pieces of information spread? And if such variations exist at the level of the underlying mechanics, can differences in the type or topic of the information help explain them?

The present work: Variation in the spread of hashtags. In this paper we analyze sources of variation in how the most widely-used hashtags on Twitter spread within its user population. We find that these sources of variation involve not just differences in the probability with which something spreads from one person to another — the quantitative analogue of stickiness — but also differences in a quantity that can be viewed as a kind of "persistence," the relative extent to which repeated exposures to a piece of information continue to have significant marginal effects on its adoption.

Moreover, these variations are aligned with the topic of the hashtag. For example, we find that hashtags on politically controversial topics are particularly persistent, with repeated exposures continuing to have large relative effects on adoption; this provides, to our knowledge, the first large-scale validation of the "complex contagion" principle from sociology, which posits that repeated exposures to an idea are particularly crucial when the idea is in some way controversial or contentious [5, 6].

Our data is drawn from a large snapshot of Twitter containing large coverage of all tweets during a period of multiple months.

WWW 2011, March 28–April 1, 2011, Hyderabad, India.
ACM 978-1-4503-0632-4/11/03.

From this dataset, we build a network on the users from the structure of interaction via @-messages; for users X and Y, if X includes "@Y" in at least t tweets, for some threshold t, we include a directed edge from X to Y. @-messages are used on Twitter for a combination of communication and name-invocation (such as mentioning a celebrity via @, even when there is no expectation that they will read the message); under all these modalities, they provide evidence that X is paying attention to Y, and with a strength that can be tuned via the parameter t.[1]

For a given user X, we call the set of other users to whom X has an edge the *neighbor set* of X. As users in X's neighbor set each mention a given hashtag H in a tweet for the first time, we look at the probability that X will first mention it as well; in effect, we are asking, "How do successive exposures to H affect the probability that X will begin mentioning it?" Concretely, following the methodology of [7], we look at all users X who have not yet mentioned H, but for whom k neighbors have; we define $p(k)$ to be the fraction of such users who mention H before a $(k+1)^{\mathrm{st}}$ neighbor does so. In other words, $p(k)$ is the fraction of users who adopt the hashtag directly after their k^{th} "exposure" to it, given that they hadn't yet adopted it.

As an example, Figure 1 shows a plot of $p(k)$ as a function of k averaged over the 500 most-mentioned hashtags in our dataset. Note that these top hashtags are used in sufficient volume that one can also construct meaningful $p(k)$ curves for each of them separately, a fact that will be important for our subsequent analysis. For now, however, we can already observe two basic features of the average $p(k)$ curve's shape: a ramp-up to a peak value that is reached relatively early (at $k = 2, 3, 4$), followed by a decline for larger values of k. In keeping with the informal discussion above, we define the *stickiness* of the curve to be the maximum value of $p(k)$ (since this is the maximum probability with which an exposure to H transfers to another user), and the *persistence* of the curve to be a measure of its rate of decay after the peak.[2] We will find that, in a precise sense, these two quantities — stickiness and persistence — are sufficient to approximately characterize the shapes of individual $p(k)$ curves.

Variation in Adoption Dynamics Across Topics. The shape of $p(k)$ averaged over all hashtags is similar to analogous curves measured recently in other domains [7], and our interest here is in going beyond this aggregate shape and understanding how these curves vary across different kinds of hashtags. To do this, we first classified the 500 most-mentioned hashtags according to their topic. We then average the curves $p(k)$ separately within each category and compare their shapes.[3]

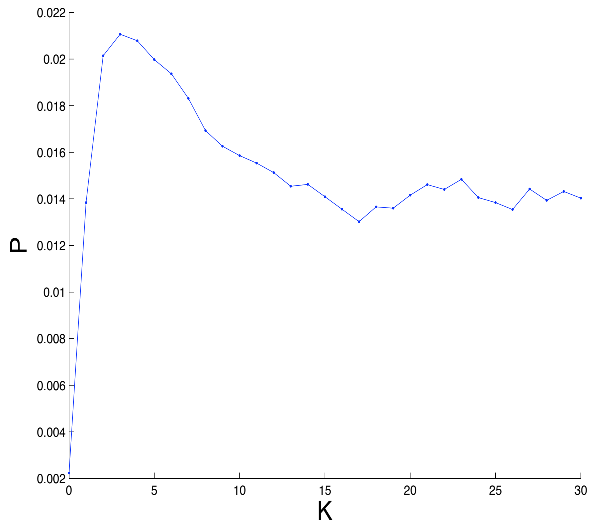


Figure 1: Average exposure curve for the top 500 hashtags. $P(K)$ is the fraction of users who adopt the hashtag directly after their k^{th} exposure to it, given that they had not yet adopted it

Many of the categories have $p(k)$ curves that do not differ significantly in shape from the average, but we find unusual shapes for several important categories. First, for political hashtags, the persistence has a significantly larger value than the average — in other words, successive exposures to a political hashtag have an unusually large effect relative to the peak. This is striking in the way that it accords with the "complex contagion" principle discussed earlier: when a particular behavior is controversial or contentious, people may need more exposure to it from others before adopting it themselves [5, 6].

In contrast, we find a different form of unusual behavior from a class of hashtags that we refer to as Twitter *idioms* — a kind of hashtag that will be familiar to Twitter users in which common English words are concatenated together to serve as a marker for a conversational theme (e.g. #cantlivewithout, #dontyouhate, #iloveitwhen, and many others, including concatenated markers for weekly Twitter events such as #musicmonday and #followfriday.) Here the stickiness is high, but the persistence is unusually low; if a user doesn't adopt an idiom after a small number of exposures, the marginal chance they do so later falls off quickly.

Subgraph Structure and Tie Strength. In addition to the person-to-person mechanics of spread, it is also interesting to look at the overall structure of interconnections among the initial adopters of a hashtag. To do this, we take the first m individuals to mention a particular hashtag H, and we study the structure of the subgraph G_m induced on these first m mentioners. In this structural context, we again find that political hashtags exhibit distinctive features — in particular, the subgraphs G_m for political hashtags H tend to exhibit higher internal degree, a greater density of triangles, and a large of number of nodes not in G_m who have significant

[1]One can also construct a directed network from the *follower* relationship, including an edge from X to Y if X follows Y. We focus here on @-messages in part because of a data resolution issues — they can be recovered with exact time stamps from the tweets themselves — but also because of earlier research suggesting that users often follow other users in huge numbers and hence potentially less discriminately, whereas interaction via @-messages indicates a kind of attention that is allocated more parsimoniously, and with a strength that can be measured by the number of repeat occurrences [17].

[2]We formally define persistence in Section 3; roughly, it is the ratio of the area under the curve to the area of the largest rectangle that can be circumscribed around it.

[3]In Section 2 we describe the methodology used to perform this manual classification in detail. In brief, we compared independent classifications of the hashtags obtained by disjoint means, involving annotation by the authors compared with independent annotation by a group of volunteers. Our results based on the average curves arising from this classification are robust in the following sense: despite differences in classification of some individual hashtags by the two groups, the curves themselves exhibit essentially identical behavior when computed from either of the two classifications separately, as well as from an intersection of the two classifications.

numbers of neighbors in it. This is again broadly consistent with the sociological premises of complex contagion, which argues that the successful spread of controversial behaviors requires a network structure with significant connectivity and significant local clustering.

Within these subgraphs, we can consider a set of sociological principles that are related to complex contagion but distinct from it, centered on the issue of *tie strength*. Work of McAdam and others has argued that the sets of early adopters of controversial or risky behaviors tend to be rich in strong ties, and that strong ties are crucial for these activities [25, 26] — in contrast to the ways in which learning about novel information can correspondingly benefit from transmission across weaker ties [13].

When we look at tie strength in these subgraphs, we find a somewhat complex picture. Because subgraphs G_m for political hashtags have significantly more edges, they have more ties of all strengths, including strong ties (according to several different definitions of strength summarized in Section 4). This aspect of the data aligns with the theories of McAdam and others. However, the fraction of strong ties in political subgraphs G_m is actually *lower* than the fraction of strong ties for the full population of widely-used hashtags, indicating the overall greater density of edges in political subgraphs comes more dominantly from a growth in weak ties than from strong ones. The picture that emerges of early-adopter subgraphs for political hashtags is thus a subtle one: they are structures whose communication patterns are more densely connected than the early-adopter subgraphs for other hashtags, and this connectivity comes from a core of strong ties embedded in an even larger profusion of weak ties.

Interpreting the Findings. When we look at politically controversial topics on Twitter, we therefore see both direct reflections and unexpected variations on the sociological theories concerning how such topics spread. This is part of a broader and important issue: understanding differences in the dynamics of contentious behavior in the off-line world versus the on-line world. It goes without saying that the use of a hashtag on Twitter isn't in any sense comparable, in terms of commitment or personal risk, to taking part in activism in the physical world (a point recently stressed in a much-circulated article by Malcolm Gladwell [12]). But the underlying issue persists on Twitter: political hashtags are still riskier to use than conversational idioms, albeit at these much lower stakes, since they involve publicly aligning yourself with a position that might alienate you from others in your social circle. The fact that we see fundamental aspects of the same sociological principles at work both on-line and off-line suggests a certain robustness to these principles, and the differences that we see suggest a perspective for developing deeper insights into the relationship between these behaviors in the on-line and off-line domains.

This distinction between contentious topics in the on-line and off-line worlds is one issue to keep in mind when interpreting these results. Another is the cumulative nature of the findings. As with any analysis at this scale, we are not focusing on why any one individual made the decisions they did, nor is it the case that that Twitter users are even aware of all the tweets containing their exposures to hashtags via neighbors. Rather, the point is that we still find a strong signal in an aggregate sense — as a whole, the population is exhibiting differences in how it responds to hashtags of different types, and in ways that accord with theoretical work in other domains.

A further point to emphasize is that our focus in this work is on the hashtags that succeeded in reaching large numbers of people. It is an interesting question to consider what distinguishes a hashtag that spreads widely from one that fails to attract attention, but that is not the central question we consider here. Rather, what we are identifying is that among hashtags that do reach many people, there can nevertheless be quite different mechanisms of contagion at work, based on variations in stickiness and persistence, and that these variations align in interesting ways with the topic of the hashtag itself.

Simulated Spreading. Finally, an interesting issue here is the interaction between the $p(k)$ curve and the subgraph G_m for a given hashtag H — clearly the two develop in a form of co-evolution, since the addition of members via the curve $p(k)$ determines how the subgraph of adopters takes shape, but the structure of this subgraph — particularly in the connections between adopters and non-adopters — affects who is likely to use the hashtag next. To understand how $p(k)$ and G_m relate to each other, it is natural to consider questions of the following form: how would the evolution of G_m have turned out differently if a different $p(k)$ curve had been in effect? Or correspondingly, how effectively would a hashtag with curve $p(k)$ have spread if it had started from a different subgraph G_m? Clearly it is difficult to directly perform this counterfactual experiment as stated, but we obtain insight into the structure of the question by simulating the $p(k)$ curve of each top hashtag on the subgraph G_m of each other top hashtag. In this way, we begin to identify some of the structural factors at work in the interplay between the mechanics of person-to-person influence and the network on which it is spreading.

2. DATASET, NETWORK DEFINITION, AND HASHTAG CLASSIFICATION

Data Collection and Network Definition. From August 2009 until January 2010 we crawled Twitter using their publicly available API. Twitter provides access to only a limited history of tweets through the search mechanism; however, because user identifiers have assigned contiguously since an early point in time, we simply crawled each user in this range. Due to limitations of the API, if a user has more than 3,200 tweets we can only recover the last 3,200 tweets; all messages of any user with fewer than this many tweets are available. We collected over three billion messages from more than 60 million users during this crawl.

As discussed in Section 1, in addition to extracting tweets and hashtags within them, we also build a network on the users, connecting user X to user Y if X directed at least t @-messages to Y. In our analyses we use $t = 3$, except when we are explicitly varying this parameter. The resulting network contains 8,509,140 non-isolated nodes and 50,814,366 links. As noted earlier, there are multiple ways of defining a network on which hashtags can viewed as diffusing, and our definition is one way of defining a proxy for the attention that users X pay to other users Y.

Hashtag Selection and Classification. To create a classification of hashtags by category, we began with the 500 hashtags in the data that had been mentioned by the most users. From manual inspection of this list, we identified eight broad categories of hashtags that each had at least 20 clear exemplars among these top hashtags, and in most cases significantly more. (Of course, many of the top 500 hashtags fit into none of the categories.) We formulated definitions of these categories as shown in Table 1. Then we applied multiple independent mechanisms for classifying the hashtags according to these categories. First, the authors independently annotated each hashtag, and then had a reconciliation phase in which

Category	Definition
Celebrity	The name of a person or group (e.g. music group) that is featured prominently in entertainment news. Political figures or commentators with a primarily political focus are not included. The name of the celebrity may be embedded in a longer hashtag referring to some event or fan group that involves the celebrity. Note that many music groups have unusual names; these still count under the "celebrity" category.
Games	Names of computer, video, MMORPG, or twitter-based games, as well as groups devoted to such games.
Idiom	A tag representing a conversational theme on twitter, consisting of a concatenation of at least two common words. The concatenation can't include names of people or places, and the full phrase can't be a proper noun in itself (e.g. a title of a song/movie/organization). Names of days are allowed in the concatenation, because of the the Twitter convention of forming hashtags involving names of days (e.g. MusicMonday). Abbreviations are allowed only if the full form also appears as a top hashtag (so this rules out hashtags including omg, wtf, lol, nsfw).
Movies/TV	Names of movies or TV shows, movie or TV studios, events involving a particular movie or TV show, or names of performers who have a movie or TV show specifically based around them. Names of people who have simply appeared on TV or in a movie do not count.
Music	Names of songs, albums, groups, movies or TV shows based around music, technology designed for playing music, or events involving any of these. Note that many music groups have unusual names; these still count under the "music" category.
Political	A hashtag that in your opinion often refers to a politically controversial topic. This can include a political figure, a political commentator, a political party or movement, a group on twitter devoted to discussing a political cause, a location in the world that is the subject of controversial political discussion, or a topic or issue that is the subject of controversial political discussion. Note that this can include political hashtags oriented around countries other than the U.S.
Sports	Names of sports teams, leagues, athletes, particular sports or sporting events, fan groups devoted to sports, or references to news items specifically involving sports.
Technology	Names of Web sites, applications, devices, or events specifically involving any of these.

Table 1: Definitions of categories used for annotation.

Category	Examples	Category	Examples
Celebrity	mj, brazilwantsjb, regis, iwantpeterfacinelli	Music	thisiswar, mj, musicmonday, pandora
Games	mafiawars, spymaster, mw2, zyngapirates	Political	tcot, glennbeck, obama, hcr
Idiom	cantlivewithout, dontyouhate, musicmonday	Sports	golf, yankees, nhl, cricket
Movies/TV	lost, glennbeck, bones, newmoon	Technology	digg, iphone, jquery, photoshop

Table 2: A small set of examples of members in each category.

they noted errors and arrived at a majority judgment on each annotation. Second, the authors solicited a group of independent annotators, and took the majority among their judgments. Annotaters were provided with the category definitions, and for each hashtag were provided with the tag's definitions (when present) from the Web resources Wthashtag and Tagalus, as well as links to Google and Twitter search results on the tag. Finally, since the definition of the "idiom" category is purely syntactic, we did not use annotators for this task, but only for the other seven categories.

Clearly even with this level of specificity, involving both human annotation and Web-based definitional resources, there are ultimately subjective judgments involved in category assignments. However, given the goal of understanding variations in hashtag behavior across topical categories, at some point in the process a set of judgments of this form is unavoidable. What we find is the results are robust in the presence of these judgments: the level of agreement among annotators was uniformly high, and the plots presented in the subsequent sections show essentially identical behavior regardless of whether they are based on the authors' annotations, the independent volunteers' annotations, or the intersection of the two. To provide the reader with some intuition for the kinds of hashtags that fit each category, we present a handful of illustrative examples in Table 2, drawn from the much larger full membership in each category. The full category memberships can be seen at http://www.cam.cornell.edu/$\sim$dromero/top500ht.

3. EXPOSURE CURVES

Basic definitions. In order to investigate the mechanisms by which hashtag usage spreads among Twitter users, we begin by reviewing two ways of measuring the impact that exposure to others has in an individual's' choice to adopt a new behavior (in this case, using a hashtag) [7]. We say that a user is $k-$*exposed* to hashtag h if he has not used h, but has edges to k other users who have used h in the past. Given a user u that is $k-$exposed to h we would like to estimate the probability that u will use h in the future. Here are two basic ways of doing this.

Ordinal time estimate. Assume that user u is $k-$exposed to some hashtag h. We will estimate the probability that u will use h before becoming $(k+1)-$exposed. Let $E(k)$ be the number of users who were $k-$exposed to h at some time, and let $I(k)$ be the number of users that were $k-$exposed and used h before becoming $(k+1)-$exposed. We then conclude that the probability of using the hashtag h while being $k-$exposed to h is $p(k) = \frac{I(k)}{E(k)}$.

Snapshot estimate. Given a time interval $T = (t_1, t_2)$, assume that a user u is $k-$exposed to some hashtag h at time $t = t_1$. We will estimate the probability that u will use h sometime during time interval T. We let $E(k)$ be the number of users who were $k-$exposed to h at time $t = t_1$, and let $I(k)$ be the number of users who were $k-$exposed to h at time $t = t_1$ and used h sometime before $t = t_2$. We then conclude that $p(k) = \frac{I(k)}{E(k)}$ is the probability of using h before time $t = t_2$, conditioned on being $k-$exposed to h at time $t = t_1$. We will refer to $p(k)$ as an *exposure curve*; we will also informally refer to it as an *influence curve*, although it is being used only for prediction, not necessarily to infer causal influence.

The ordinal time approach requires more detailed data than the snapshot method. Since our data are detailed enough that we are able to generate the ordinal time estimate, we only present the results based on the ordinal time approach; however, we have confirmed that the conclusions hold regardless of which approached is followed. This is not surprising since it has been argued that sufficiently many snapshot estimates contain enough information to infer the the ordinal time estimate [7].

Comparison of Hashtag Categories: Persistence and Stickiness. We calculated ordinal time estimates $P(k)$ for each one of the 500 hashtags we consider. For each point on each curve we calculate the 95% Binomial proportion confidence interval. We observed some qualitative differences between the curves corresponding to different hashtags. In particular, we noticed that some curves increased dramatically initially as k increased but then started to decrease relatively fast, while other curves increased at a much slower rate initially but then saturated or decreased at a much slower rate. As an example, Figure 3 shows the influence curves for the hashtags

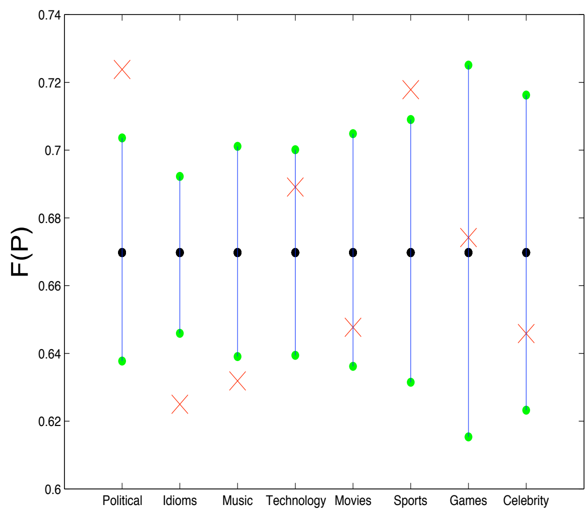


Figure 2: $F(P)$ for the different types of hashtags.The black dots are the average $F(P)$ among all hashtags, the red x is the average for the specific category, and the green dots indicate the 90% expected interval where the average for the specific set of hashtags would be if the set was chosen at random. Each point is the average of a set of at least 10 hashtags

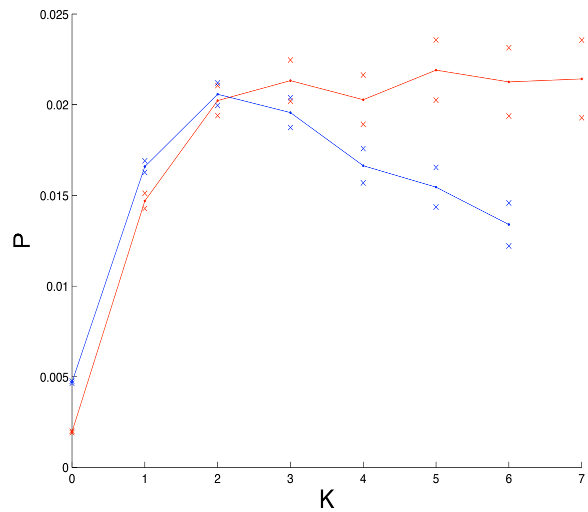


Figure 3: Sample exposure curves for hashtags #cantlivewithout (blue) and #hcr (red).

#cantlivewithout and #hcr. We also noticed that some curves had much higher maximum values than others.[4]

In this discussion, we are basing differences among hashtags on different structural properties of their influence curves. In order to make these distinctions more precise we use the following measures.

First, we formalize a notion of "persistence" for an influence curve, capturing how rapidly it decays. Formally, given a function $P : [0, K] \to [0, 1]$ we let $R(P) = K \max_{k\in[0,K]}\{P(k)\}$ be the area of the rectangle with length K and height $\max_{k\in[0,K]}\{P(k)\}$. We let $A(P)$ be the area under the curve P assuming the point $P(k)$ is connected to the point $P(k + 1)$ by a straight line. Finally, we let $F(P) = \dfrac{A(P)}{R(P)}$ be the *persistence* parameter.

When an influence curve P initially increases rapidly and then decreases, it will have a smaller value of $F(P)$ than a curve $\widetilde{P}$ which increases slowly and the saturates. Similarly, an influence curve P that slowly increases monotonically will have a smaller value of $F(P)$ than a curve $\widetilde{P}$ that initially increases rapidly and then saturates. Hence the measure F captures some differences in the shapes of the influence curves. In particular, applying this measure to an influence curve would tell us something about its persistence; the higher the value of $F(P)$, the more persistent P is.

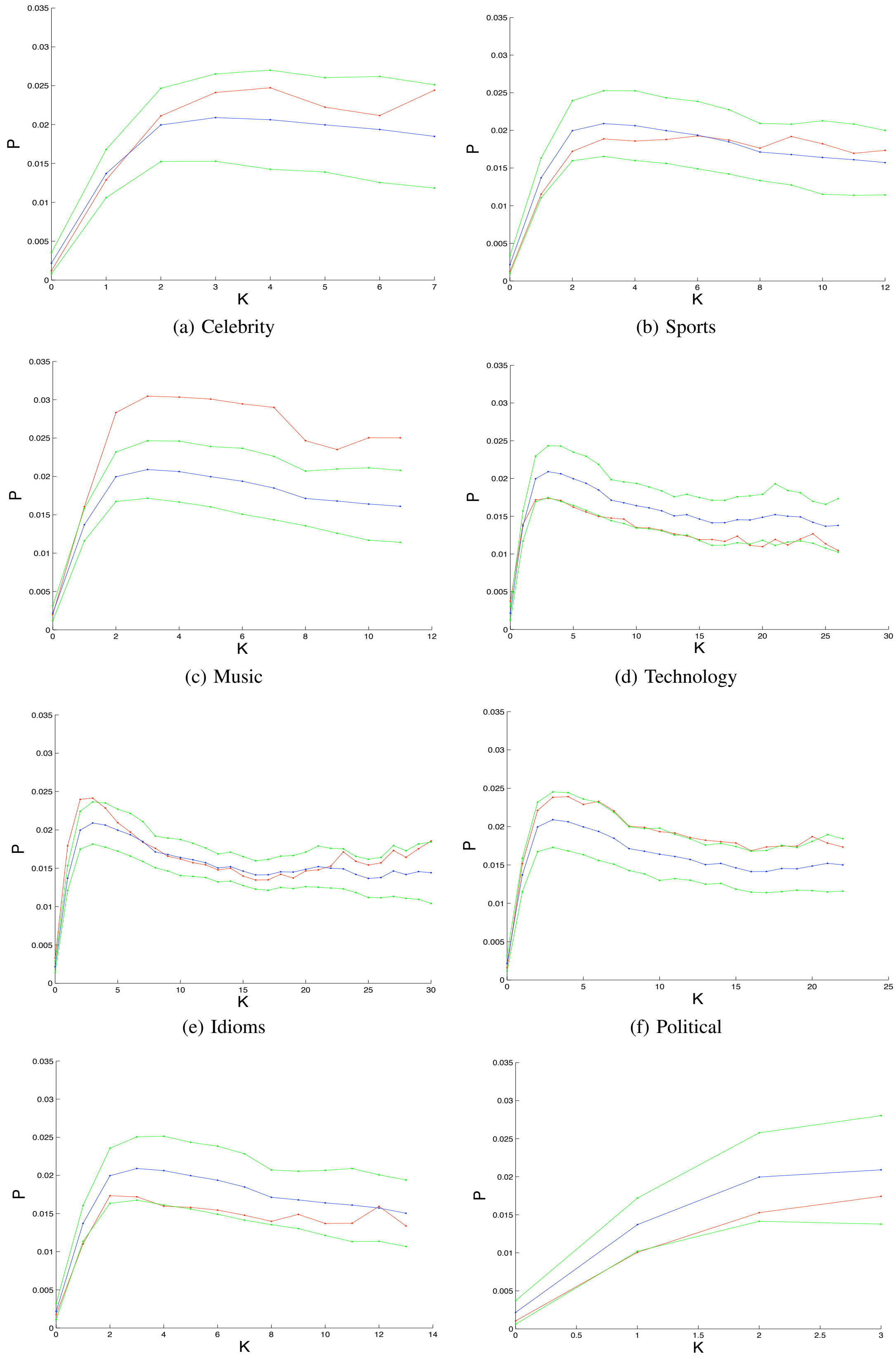
Second, given an influence curve $P : [0, K] \to [0, 1]$ we let $M(P) = \max_{k\in[0,K]}\{P(k)\}$ be the *stickiness* parameter, which gives us a sense for how large the probability of usage can be for a particular hashtag based on the most effective exposure.

[4]As k gets larger the amount of data used to calculate $P(k)$ decreases, making the error intervals very large and the curve very noisy. In order to take this into account we only defined $P(k)$ when the relative error was less than some value θ. Throughout the study we checked that the results held for different values of θ.

We are interested in finding differences between the spreading mechanism of different topics on Twitter. We start by finding out if hashtags corresponding to different topics have influence curves with different shapes. We found significant differences in the values of $F(P)$ for different topics. Figure 2 shows the average $F(P)$ for the different categories, compared to a baseline in which we draw a set of categories of the same size uniformly at random from the full collection of 500. We see that politics and sports have an average value of $F(P)$ which is significantly higher than expected by chance, while for Idioms and Music it is lower. This suggests that the mechanism that controls the spread of hashtags related to sports or politics tends to be more persistent than average; repeated exposures to users who use these hashtags affects the probability that a person will eventually use the hashtag more positively than average. On the other hand, for Idioms and Music, the effect of repeated exposures falls off more quickly, relative to the peak, compared to average.

Figure 4 shows the point-wise average of the influence curves for each one of the categories. Here we can see some of the differences in persistence and stickiness the curves have. For example, the stickiness of the topics Music, Celebrity, Idioms, and politics tends to be higher that average since the average influence curve for those categories tends to be higher than the average influence curve for all hashtags, while that of Technology, Movies, and Sports tends to be lower than average. On the other hand, these plots give us more intuition on why we found that politics and Sports have a high persistence while for Idioms and Music it is low. In the case of Politics, we see that the red curve starts off just below the green curve (the upper error bar) and as k increases, the red curve increases enough to be above the green. Similarly, the red curve for Sports starts below the blue curve and it ends above it. In the case of Idioms, the red curve initially increases rapidly but then it it drops below the blue curve. Similarly, the red curve for Music is always very high and above all the other curves, but it drops faster than the other curves at the end.

Approximating Curves via Stickiness and Persistence. When we compare curves based on their stickiness and persistence, it



(a) Celebrity

(b) Sports

(c) Music

(d) Technology

(e) Idioms

(f) Political

(g) Movies

(h) Games

Figure 4: Point-wise average influence curves. The blue line is the average of all the influence curves, the red line is the average for the set of hashtags of the particular topic, and the green lines indicate the interval where the red line is expected to be if the hashtags were chosen at random.

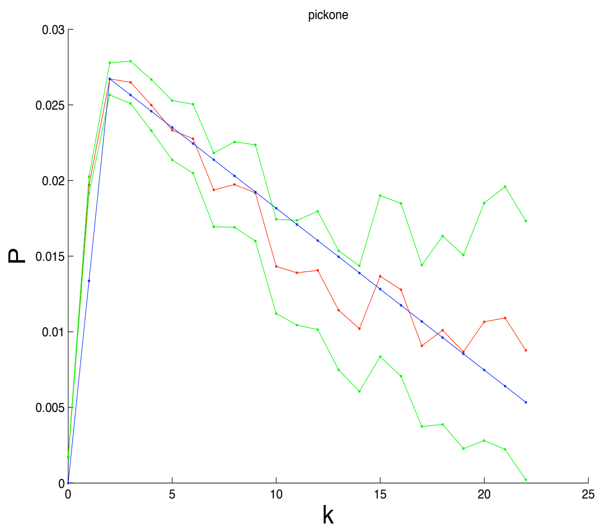


Figure 5: Example of the approximation of an influence curve. The red curve is the influence curve for the hashtag #pickone, the green curves indicate the 95% binomial confidence interval, and the blue curve is the approximation.

is important to ask whether these are indeed an adequate pair of parameters for discussing the curves' overall "shapes." We now establish that they are, in the following sense: we show that these two parameters capture enough information about the influence curves that we can approximate the curves reasonably well given just these two parameters. Assume that for some curve P we are given $F(P)$ and $M(P)$. We will also assume that we know the maximum value of $k = K$ for which $P(k)$ is defined. Then we will construct an approximation curve $\widetilde{P}$ in the following way:

1. Let $\widetilde{P}(0) = 0$
2. Let $\widetilde{P}(2) = M(P)$
3. Now we will let $\widetilde{P}(K)$ be such that $F(\widetilde{P}) = F(P)$. This value turns out to be $\widetilde{P}(K) = \frac{M(P)*K*(2*F(P)-1)}{K-2}$
4. Finally, we will make $\widetilde{P}$ be piecewise linear with one line connecting the points $(0, 0)$ and $(2, M(P))$, and another line connecting the points $(2, M(P))$ and $(K, \frac{M(P)*K*(2*F(P)-1)}{K-2})$.

Figure 5 shows an example of an approximation for a particular influence curve. In order to test the quality of the approximation $\widetilde{P}$ we define the approximation error between $\widetilde{P}$ and P as the mean absolute error

$$E(P, \widetilde{P}) = \frac{1}{K}\sum_{k=0}^{K}\left|(P(k) - \widetilde{P}(k))\right|$$

and compare it with the mean absolute of the error $E(P)$ obtained from the 95% confidence intervals around each point $P(k)$. The average approximation error among all the influence curves is 0.0056 and the average error of based on the confidence intervals is 0.0050. The approximation error is slightly smaller, which means that out approximation is, on average, within the 95% confidence interval from the actual influence curve. This suggests the information contained in the stickiness and persistence parameters are enough to accurately approximate the influence curves and gives more meaning to the approach of comparing the curves by comparing these two parameters.

Type	Mdn. Mentions	Mdn. Users	Mdn. Ment./User
All HTS	93,056	15,418	6.59
Political	132,180	13,739	10.17
Sports	98,234	11,329	9.97
Idioms	99,317	26,319	3.54
Movies	90,425	15,957	6.57
Celebrity	87,653	5,351	17.68
Technology	90,462	24,648	5.08
Games	123,508	15,325	6.61
Music	87,985	7,976	10.39

Table 3: Median values for number of mentions, number of users, and number of mentions per user for different types of hashtags

Frequency of Hashtag Usage. We have observed that different topics have differences in their spreading mechanisms. We also found that they differed in other ways. For example, we see some variation in the number of mentions and the number of users of each category. Table 3 shows the different median values for number of mentions, number of users, and number of mentions per user for different types of hashtags. We see that while Idioms and Technology hashtags are used by many users compared to others, each user only uses the hashtag a few times and hence the total number of mentions of the these categories is not much higher than others. On the other hand, only relatively few people used Political and Games hashtags, but each one of them used them many times, making them the most mentioned categories. In the case of games, a contributing factor is that some of users of game hashtags allow external websites to post on their Twitter account every time they accomplish something in the game, which tends to happen very often. It is not clear that there is a correspondingly simple explanation for the large number of mentions per user for political hashtags, but one can certainly conjecture that it may reflect something about the intensity with which these topics are discussed by the users who engage in such discussions; this is an interesting issue to explore further.

4. THE STRUCTURE OF INITIAL SETS

The spread of a given piece of information is affected by the diffusion mechanism controlled by the influence curves discussed in the previous section, but it may also be affected by the structure of the network relative to the users of the hashtag. To explore this further, we looked at the subgraph G_m induced by the first m people who used a given hashtag. We found that there are important differences in the structure of those graphs.

In particular, we consider differences in the structures of the subgraphs G_m across different categories. For each graph G_m, across all hashtags and a sequence of values of m, we compute several structural parameters. First, we compute the average degree of the nodes and the number of triangles in the graph. Then, we defined the *border* of G_m to be the set of all nodes not in G_m who have at least one edge to a node in G_m, and we define the *entering degree* of a node in the border to be the number of neighbors it has in G_m. We consider the size of the border and the average entering degree of nodes in the border.

Looking across all categories, we find that political hashtags are

Type	I	II	III	IV
All HTS	1.41	384	1.24	13425
Political	2.55	935	1.41	12879
Upper Error Bar	1.82	653	1.32	15838
Lower Error Bar	1.00	112	1.16	11016

Table 4: Comparison of graphs induced by the first 500 early adopters of political hashtags and average hashtags. Column definitions: I. Average degree, II. Average triangle count, III. Average entering degree of the nodes in the border of the graphs, IV. Average number of nodes in the border of the graphs. The error bars indicate the 95% confidence interval of the average value of a randomly selected set of hashtags of the same size as Political.

the category in which the most significant structural differences from the average occur. Table 4 shows the averages for political hashtags compared to the average for all hashtags, using the subgraphs G_{500} on the first 500 users.[5] In brief, the early adopters of a political hashtag message with more people, creating more triangles, and with a border of people who have more links on average into the early adopter set. The number of triangles, in fact, is high even given the high average degree; clearly one should expect a larger number of triangles in a subgraph of larger average degree, but in fact the triangle count for political hashtags is high even when compared against a baseline consisting of non-political hashtags with comparable average degrees. These large numbers of edges and triangles are consistent with the predictions of complex contagion, which argues that such structural properties are important for the spread of controversial topics [6].

Tie Strength. There is an interesting further aspect to these structural results, obtained by looking at the *strength* of the ties within these subgraphs. There are multiple ways of defining tie strength from social media data [10], and here we consider two distinct approaches. One approach is to use the total number of @-messages sent across the link as a numerical measure of strength. Alternately, we can declare a link to be strong if and only if it is *reciprocated* (i.e. declaring (X, Y) to be strong if and only if (Y, X) is in the subgraph as well, following a standard working notion of reciprocation as a proxy for tie strength in the sociology literature [14]).

Under both definitions, we find that the fraction of strong ties in subgraphs G_m for political hashtags is in fact significantly lower than the fraction of strong ties in subgraphs G_m for our set of hashtags overall. However, since political subgraphs G_m contain so many links relative to the typical G_m, we find that they have a larger absolute number of strong ties. As noted in the introduction, standard sociological theories suggest that we should see many strong ties in subgraphs G_m for political topics, but the picture we obtain is more subtle in that the growth in strong ties comes with an even more significant growth in weak ties. Understanding these competing forces in the structural behavior of such subgraphs is an interesting open question.

5. SIMULATIONS

We have observed that for some hashtags, such as those relating to political subjects, users are particularly affected by multiple exposures before using them. We also know that the subgraphs on which political hashtags initially spread have high degrees and extensive clustering. To what extent do these aspects intrinsically go together? Do these types of political hashtags spread effectively because of the close-knit network of the initial users? Are political subjects less likely to successfully spread on sparsely connected initial sets?

In this section, we try to obtain some initial insight into these questions through a simulation model — not only in the context of political hashtags but also in the context of the other categories. In particular, we develop a model that naturally complements the process used to calculate the $p(k)$ functions. We perform simulations of this model using the measured $p(k)$ functions and a varying number of the first users who used each hashtag on the actual influence network. Additionally, we record the progression of the cascade and track its spread through the network. By trying the $p(k)$ curve of a hashtag on the initial sets of other hashtags, and by varying the size of the initial sets, we can gain insight into the factors that lead to wide-spreading cascades.

5.1 The Simulated Model

We wish to simulate cascades using the measured $p(k)$ curves, the underlying network of users, and in particular the observed subgraphs G_m of initial adopters, In this discussion, and in motivating the model, we refer to the moment at which a node adopts a hashtag as its *activation*. We operationalize the model implicit in the definition of the function $p(k)$, leading to the following natural simulation process on a graph $G = (V, E)$.

First, we activate all nodes in the starting set I, and mark them all as newly active. In a general iteration t (starting with $t = 0$), we will have a currently active set A_t and a subset $N_t \subseteq A_t$ of *newly active* nodes. (In the opening iteration, we have $A_0 = N_0 = I$.) Newly active nodes have an opportunity to activate nodes $u \in V - A_t$, with the probabilities of success on u determined by the $p(k)$ curve and the number of nodes in $A_t - N_t$ who have already tried and failed to activate u.

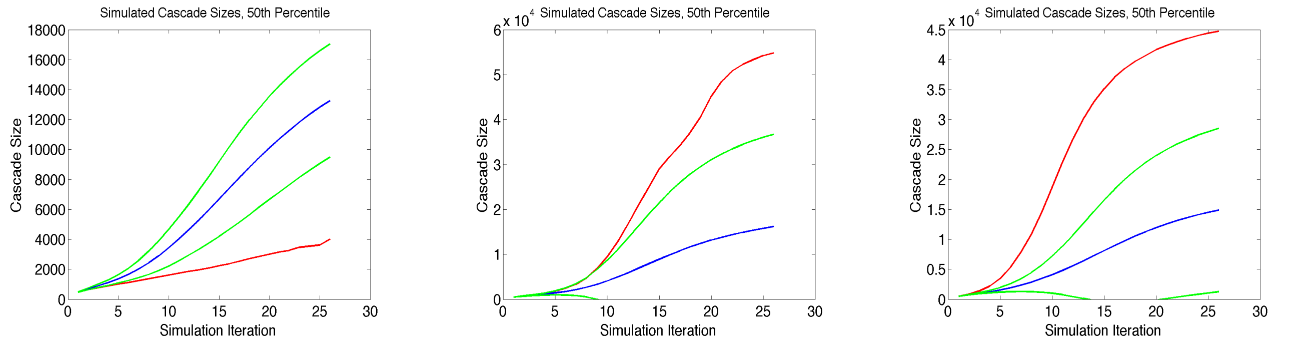
Thus, we consider each node $u \in V - A_t$ that is a neighbor of at least one node in N_t, and hence will experience at least one activation attempt. Let $k_t(u)$ be the number of nodes in $A_t - N_t$ adjacent to u; these are the nodes that have already tried and failed to activate u. Let $\Delta_t(u)$ be the number of nodes in N_t adjacent to u. Each of these neighbors in N_t will attempt to activate u in sequence, and they will succeed with probabilities $p(k_t(u) + 1), p(k_t(u) + 2), \ldots, p(k_t(u) + \Delta_t(u))$, since these are the success probabilities given the number of nodes that have already tried and failed to activate u. At the end, we define N_{t+1} to be the set of nodes u that are newly activated by the attempts in this iteration, and $A_{t+1} = A_t \cup N_{t+1}$.

5.2 Simulation Results

We simulate how a cascade that spreads according to the $p(k)$ curve for some hashtag evolves when seeded with an initially active user sets of other hashtags. In total, there are 250,000 ($p(k)$, start set) hashtag combinations we examine. We additionally vary the size of the initially active set to be 100, 500, or 1,000 users. Since we want to study how a hashtag blossoms from being used by a few starting nodes to a large number of users, we must be careful about how we select the size of our starting sets. We believe that these initial set sizes capture the varying topology observed in Section 4 and are not too large as to guarantee wide-spreading cascade. For 100 and 500 starting nodes we run five simulations on each ($p(k)$, start set) pair, and for 1,000 starting nodes we run only two simulations.

The simulation is instrumented at each iteration; we record the size of the cascade, the number of nodes influenced by active users,

[5]The results are similar for G_m with a range of other values of $m \neq 500$.



(a) Celebrity vs. random $p(k)$ curves, celebrity start sets

(b) Political vs. random start sets, political $p(k)$ curves.

(c) Idiom vs. random start sets, idiom $p(k)$ curves.

Figure 6: Validating Category Differences: The median cascade sizes for three different categories. In (a) we randomize over the $p(k)$ curves and show that celebrity $p(k)$ curves don't perform as well as random $p(k)$ curves on celebrity start sets. Figures (b) and (c) illustrate the strength of the starting sets for political and idiom hashtags compared to random start sets. All starting sets consist of 500 users.

and the number of inactive users influenced by active users. Furthermore, each simulation runs for at most 25 iterations. We found that this number of iterations was large enough to observe interesting variation in cascade sizes yet still be efficiently simulated.

We calculate the mean and the 5th, 10th, ..., 95th percentiles of cascade sizes after each iteration. For each category, we measure these twenty measures based on all of the simulations where the $p(k)$ hashtag and the starting set hashtag are both chosen from the category. We then compare these measurements to the results when a random set of hashtags is used to decide the $p(k)$ curve, the starting set, or both the $p(k)$ curve and the starting set. The cardinality of this random set is the same as the number of hashtags in the category. We sample these random choices 10,000 times to estimate the distribution of these measured features.

Using these samples, we test the measurements for statistical significance. In particular, we look at how the 'category' cascades (those in which both hashtag choices are from the category set) compare to cascades in which the $p(k)$ curve or starting set hashtages were chosen randomly. In all of the following figures, the red line indicates the value of the measurements over the set of simulations in which $p(k)$ curve and the start set come from category hashtags. The blue line is the average feature measurement over the random choices, and the green lines specify two standard deviations from the mean value. The cascade behavior of a category is statistically significant with respect to one of the measured features when most of the red curve lies outside of the region between the two green curves.

We compare how the $p(k)$ curves for a category perform on start sets from the same category and on random start sets. We additionally evaluate how random $p(k)$ curves and category $p(k)$ curves perform on category start sets. In general, categories either performed below or above the random sets in both of these measures. Some particular observations are

- Celebrities and Games: Compared to random starting sets, we find that start sets from these categories generate smaller cascades when the $p(k)$ curves are chosen from their respective categories. This difference is statistically significant.
- Political and Idioms: These categories' $p(k)$ curves and start sets perform better than a random choice. This is especially true for the smaller cascades (5 - 30th percentiles).
- Music: This category is interesting because the music $p(k)$ curves perform better than random $p(k)$ curves on music starting sets, music $p(k)$ curves perform better on random starting sets than on music starting sets, regardless of the number of initially active users. This is the only category in which the $p(k)$ and start set 'goodness' differs.
- Movies, Sports, and Technology: These categories don't exhibit particularly strong over or underperformance compared a random choice of $p(k)$ hashtags and starting set hashtags.

6. CONCLUSION

By studying the ways in which an individual's use of widely-adopted Twitter hashtags depends on the usage patterns of their network neighbors, we have found that hashtags of different types and topics exhibit different mechanics of spread. These differences can be analyzed in terms of the probabilities that users adopt a hashtag after repeated exposure to it, with variations occurring not just in the absolute magnitudes of these probabilities but also in their rate of decay. Some of the most significant differences in hashtag adoption provide intriguing confirmation of sociological theories developed in the off-line world. In particular, the adoption of politically controversial hashtags is especially affected by multiple repeated exposures, while such repeated exposures have a much less important marginal effect on the adoption of conversational idioms.

This extension of information diffusion analysis, taking into account sources of variation across topics, opens up a variety of further directions for investigation. First, the process of diffusion is well-known to be governed both by influence and also by homophily — people who are linked tend to share attributes that promote similiarities in behavior. Recent work has investigated this interplay of influence and homophily in the spreading of on-line behaviors [2, 8, 3, 19]; It would be interesting to look at how this varies across topics and categories of information as well — it is plausible, for example, that the joint mention of a political hashtag provides stronger evidence of user-to-user similarity than the analogous joint mention of hashtags on other topics, or that certain conversational idioms (those that are indicative of shared background) are significantly better indicators of similarity than others. There has also been work on the temporal patterns of information diffusion — the rate over time at which different pieces of information are adopted [9, 18, 21, 24, 30]. In this context there have been comparisons between the temporal patterns of expected versus unexpected information [9] and between different media such as news sources and blogs [21]. Our analysis here suggests that a rich spectrum of differences may exist across topics as well.

Finally, we should emphasize one of our original points, that the phenomena we are observing are clearly taking place in aggregate: it is striking that, despite the many different styles in which people use a medium like Twitter, sociological principles such as the complex contagion of controversial topics can still be observed at the population level. Ultimately, it will be interesting to pursue more fine-grained analyses as well, understanding how patterns of variation at the level of individuals contribute to the overall effects that we observe.

Acknowledgements. We thank Luis von Ahn for valuable discussions and advice about this research, Curt Meeder for helping with edits, and our volunteers Ariel Levavi, Yarun Luon, and Alicia Urdapilleta for their valuable help. This work has been supported in part by the MacArthur Foundation, a Google Research Grant, a Yahoo! Research Alliance Grant, and NSF grants IIS-0705774, IIS-0910664, IIS-0910453, and CCF-0910940. Brendan Meeder is supported by a NSF Graduate Research Fellowship.

7. REFERENCES

[1] E. Adar, L. Zhang, L. A. Adamic, and R. M. Lukose. Implicit structure and the dynamics of blogspace. In *Workshop on the Weblogging Ecosystem*, 2004.

[2] A. Anagnostopoulos, R. Kumar, and M. Mahdian. Influence and correlation in social networks. In *Proc. 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 7–15, 2008.

[3] S. Aral, L. Muchnik, and A. Sundararajan. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proc. Natl. Acad. Sci. USA*, 106(51):21544–21549, Dec. 2009.

[4] L. Backstrom, D. Huttenlocher, J. Kleinberg, and X. Lan. Group formation in large social networks: Membership, growth, and evolution. In *Proc. 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2006.

[5] D. Centola. The spread of behavior in an online social network experiment. *Science*, 329(5996):1194–1197, 3 September 2010.

[6] D. Centola and M. Macy. Complex contagions and the weakness of long ties. *American Journal of Sociology*, 113:702–734, 2007.

[7] D. Cosley, D. P. Huttenlocher, J. M. Kleinberg, X. Lan, and S. Suri. Sequential influence models in social networks. In *Proc. 4th International Conference on Weblogs and Social Media*, 2010.

[8] D. Crandall, D. Cosley, D. Huttenlocher, J. Kleinberg, and S. Suri. Feedback effects between similarity and social influence in online communities. In *Proc. 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 160–168, 2008.

[9] R. Crane and D. Sornette. Robust dynamic classes revealed by measuring the response function of a social system. *Proc. Natl. Acad. Sci. USA*, 105(41):15649–15653, 29 September 2008.

[10] E. Gilbert and K. Karahalios. Predicting tie strength with social media. In *Proc. 27th ACM Conference on Human Factors in Computing Systems*, pages 211–220, 2009.

[11] M. Gladwell. *The Tipping Point: How Little Things Can Make a Big Difference*. Little, Brown, 2000.

[12] M. Gladwell. Small change: Why the revolution will not be tweeted. *The New Yorker*, 4 October 2010.

[13] M. Granovetter. The strength of weak ties. *American Journal of Sociology*, 78:1360–1380, 1973.

[14] M. Granovetter. The strength of weak ties: A network theory revisited. *Sociological Theory*, 1:201–233, 1983.

[15] D. Gruhl, D. Liben-Nowell, R. V. Guha, and A. Tomkins. Information diffusion through blogspace. In *Proc. 13th International World Wide Web Conference*, 2004.

[16] C. Heath and D. Heath. *Made to Stick: Why Some Ideas Survive and Others Die*. Random House, 2007.

[17] B. A. Huberman, D. M. Romero, and F. Wu. Social networks that matter: Twitter under the microscope. *First Monday*, 14(1), Jan. 2009.

[18] A. Johansen. Probing human response times. *Physica A*, 338(1–2):286–291, 2004.

[19] G. Kossinets and D. Watts. Origins of homophily in an evolving social network. *American Journal of Sociology*, 115(2):405–50, Sept. 2009.

[20] J. Leskovec, L. Adamic, and B. Huberman. The dynamics of viral marketing. *ACM Transactions on the Web*, 1(1), May 2007.

[21] J. Leskovec, L. Backstrom, and J. Kleinberg. Meme-tracking and the dynamics of the news cycle. In *Proc. 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2009.

[22] J. Leskovec, M. McGlohon, C. Faloutsos, N. Glance, and M. Hurst. Cascading behavior in large blog graphs. In *Proc. SIAM International Conference on Data Mining*, 2007.

[23] D. Liben-Nowell and J. Kleinberg. Tracing information flow on a global scale using Internet chain-letter data. *Proc. Natl. Acad. Sci. USA*, 105(12):4633–4638, Mar. 2008.

[24] R. D. Malmgren, D. B. Stouffer, A. E. Motter, and L. A. N. Amaral. A poissonian explanation for heavy tails in e-mail communication. *Proc. Natl. Acad. Sci. USA*, 105(47):18153–18158, 25 November 2008.

[25] D. McAdam. Recruitment to high-risk activism: The case of Freedom Summer. *American Journal of Sociology*, 92:64–90, 1986.

[26] D. McAdam. *Freedom Summer*. Oxford University Press, 1988.

[27] E. Rogers. *Diffusion of Innovations*. Free Press, fourth edition, 1995.

[28] D. Strang and S. Soule. Diffusion in organizations and social movements: From hybrid corn to poison pills. *Annual Review of Sociology*, 24:265–290, 1998.

[29] E. Sun, I. Rosenn, C. Marlow, and T. M. Lento. Gesundheit! Modeling contagion through Facebook News Feed. In *Proc. 3rd International Conference on Weblogs and Social Media*, 2009.

[30] A. Vazquez, J. G. Oliveira, Z. Deszo, K.-I. Goh, I. Kondor, and A.-L. Barabasi. Modeling bursts and heavy tails in human dynamics. *Physical Review E*, 73(036127), 2006.