Capturing Unobserved Correlated Effects in Diffusion in Large Virtual Networks

Elenna R. Dugundji, Michiel van Meeteren
Universiteit van Amsterdam
Amsterdam, Netherlands
e.r.dugundji, michielvanmeeteren@gmail.com

Ate Poorthuis
University of Kentucky
Lexington, KY, USA
atepoorthuis@gmail.com

Abstract

Social networks and social capital are generally considered to be important variables in explaining the diffusion of behavior. However, it is contested whether the actual social connections, cultural discourse, or individual preferences determine this diffusion. Using discrete choice analysis applied to longitudinal Twitter data, we are able to distinguish between social network influence on one hand and cultural discourse and individual preferences on the other hand. In addition, we present a method using freely available software to estimate the size of the error due to unobserved correlated effects. We show that even in a seemingly saturated model, the log likelihood can increase dramatically by accounting for unobserved correlated effects. Furthermore the estimated coefficients in an uncorrected model can be significantly biased beyond standard error margins.

1 Introduction

With the onset of ubiquitous social media technology, people leave numerous traces of their social behavior in – often publicly available – data sets. In this paper we look at a virtual community of independent ("Indie") software developers for the Macintosh and iPhone that use the social networking site Twitter. Using Twitter's API, we collect longitudinal data on network connections among the Indie developers and their friends and followers (approximately 15,000 nodes) and their use of Twitter client software over a period of five weeks (more than 600,000 "tweets"). We use this dynamic data on the network and user behavior to analyze the diffusion of Twitter client software.

Within the Indie community, four prominent software developers have developed Twitter clients (Tweetie, Twitterrific, Twittelator, Birdfeed) that compete for adoption within the community. Apart from these Indie Twitter clients, members of the virtual community can choose from a range of clients that are developed outside of the Indie community (for example, Tweetdeck, Twitterfon) as well as the standard Web interface provided by Twitter. Previous qualitative ethnographic evidence for our case study indicates that social networks and social capital are considered to be important factors in explaining the adoption and diffusion of behavior. [1] Using discrete choice analysis applied to longitudinal panel data, we are able to quantitatively test for the relative importance of global cultural discourse, taste-maker influence and other contextual effects, node level behavioral characteristics, socio-centric network measures and ego-centric network measures, individual preferences and social network contagion, in users' decisions of what client software they choose to interface to Twitter.

Importantly, we furthermore demonstrate a method using readily available software to estimate the size of the error due to unobserved correlated effects in users' choices. This is critical to test for in any application of multinomial logistic regression where social influence variables and/or other network measures are used as explanatory variables, since their use poses a classic case of endogeneity. We show that even in a seemingly saturated model, the log likelihood of the model fit can increase significantly by accounting for unobserved correlated effects. Furthermore the estimated coefficients in the uncorrected model can be significantly biased beyond standard error margins. Failing to account for correlated effects can yield misleading market share predictions for users' preferences for Twitter clients.

The paper is organized as follows. First a brief review of literature is presented describing what the paper brings to an existing stream of behavioral modeling research. Next the understanding of the context of the case study and the insights from the available data lead us to define nine sets of different kinds of social and individual explanatory variables to explore in our model, with different functional forms. Estimation results are summarized. Finally, directions for future research efforts are outlined.

2 Discrete choice with social interactions

2.1 Multinomial logit model

Discrete choice analysis allows prediction based on computed individual choice probabilities for heterogeneous agents' evaluation of alternatives. In accordance with notation and convention in Ben-Akiva and Lerman [2], the multinomial logit model is specified as follows. Assume a sample of N decision-making entities indexed (1,...,*n*,...,*N*) each faced with a choice among J_n alternatives indexed (1,...,*j*,..., J_n) in subset C_n of some universal choice set *C*.

The choice alternatives are assumed to be mutually exclusive (a choice for one alternative excludes the simultaneous choice for another alternative, that is, an agent cannot choose two alternatives at the same moment in time) and collectively exhaustive within C_n (an agent must make a choice for one of the options in the agent's choice set). In general the composite choice set C_n will vary in size and content across agents: not all elemental alternatives in the universal choice set may be available to all agents. For simplicity in this paper however, we will assume that the choices are available to all agents.

Let $U_{in} = V_{in} + \varepsilon_n$ be the utility that a given decision-making entity *n* is presumed to associate with a particular alternative *i* in its choice set C_n, where V_{in} is the deterministic (to the modeler) or so-called "systematic" utility and ε_{in} is an error term. Then, under the assumption of independent and identically Gumbel distributed disturbances ε_{in}, the probability that the individual decision-making entity *n* chooses alternative *i* within the choice set C_n is given by:

$$\begin{aligned} P_{in} &\equiv P_n(i \mid C_n) = \Pr\left(V_{in} + \varepsilon_{in} \geq V_{jn} + \varepsilon_{jn}, \forall j \in C_n\right) \\ &= \Pr\left[V_{in} + \varepsilon_{in} \geq \max_{j \in C_n}\left(V_{jn} + \varepsilon_{jn}\right)\right] = \frac{e^{\mu V_{in}}}{\sum_{\forall j \in C_n} e^{\mu V_{jn}}} \end{aligned}$$

where μ is a strictly positive scale parameter which is typically normalized to 1 in the multinomial logit model.

The systematic utility is commonly assumed to be defined by a linear-in-parameters function of observable characteristics $\mathbf{S}_n$ of the decision-making entity and observable attributes $\mathbf{z}_{in}$ of the choice alternative for a given decision-making entity:

$$V_{in} = h_i + V\left(\mathbf{S}_n, \mathbf{z}_{in}\right) = h_i + \boldsymbol{\gamma}_i{}'\mathbf{S}_n + \boldsymbol{\zeta}_i{}'\mathbf{z}_{in}$$

The term h_i is a so-called "alternative specific constant" (ASC), as good practice to explicitly account for any underlying bias for one alternative over another alternative. In other words, h_i reflects the mean of $\varepsilon_{jn} - \varepsilon_{in}$, that is, the difference in the utility of alternative *i* from that of *j* when all else is equal. Since it is the difference that is relevant, for a general multinomial case with J alternatives we can define a set of at most $J – 1$ alternative specific constants.

The terms $\boldsymbol{\gamma}_i = [\gamma_{i1}, \gamma_{i2}, \ldots]'$ and $\boldsymbol{\zeta}_i = [\zeta_{i1}, \zeta_{i2}, \ldots]'$ are vectors of unknown utility parameters respectively corresponding to the relevant observable agent characteristics $\mathbf{S}_n$, and observable agent-specific attributes $\mathbf{z}_{in}$ of the choice alternatives. In general the utility parameters may take alternative specific values, however when there is no variation of the agent characteristics $\mathbf{S}_n$ across the choice alternatives, we can define a set of at most $J – 1$ vectors of alternative specific coefficients for the case of the $\boldsymbol{\gamma}_i$.

2.2 Social interactions

An outstanding challenge in discrete choice analysis is the treatment of the interdependence of various decision-makers' choices [3,4]. Brock and Durlauf [5] introduce social interactions in multinomial discrete choice models by allowing a given agent's choice for a particular alternative to be dependent on the overall share of decision makers who choose that alternative. If the coefficient on this interaction variable is close to zero and not important relative to other contributions to the utility, then the distribution of decision-makers' choices will not effectively change over time in relation to other decision-makers' choices. However, if the coefficient on this interaction variable is positive and dominant enough relative to other contributions to utility, there may arise a runaway situation over time as all decision-makers flock to one particularly attractive choice alternative. In short, the specification captures social feedback between decision-makers that can potentially be reinforcing over the course of time. In diverse literature this is referred to as a social multiplier, a cascade, a bandwagon effect, imitation, contagion, herd behavior, etc. [6]

We introduce a social feedback effect among agents by allowing the systematic utility Vin to be a linear-in-parameter β first-order function of the proportion x_{in} of a given decision-maker's reference entities who have made this choice. Our model differs from the Brock and Durlauf model in that we consider *non-global* interactions. Agents see different proportions, depending on who their particular reference entities are. Additionally, we also consider various socio-centric and ego-centric network measures and other explanatory variables as contributions to the utility.

2.3 Endogeneity

One econometric issue that arises in empirical estimation of social interactions in discrete choice models using standard multinomial logistic regression however, is that the error terms are assumed to be identically and independently distributed across decision-makers. It is not obvious that this is in fact a valid assumption when we are specifically considering interdependence between decision-makers' choices. We might reason that if there is a systematic dependence of each decision-maker's choice on an explanatory variable that captures the aggregate choices of other decision-makers who are in some way related to that decision-maker, then there might be an analogous dependence in the error structure. Otherwise said, the same unobserved effects might be likely to influence the choice made by a given decision-maker as well as the choices made by those in the decision-maker's reference group, which is a classic case of endogeneity. The results and coefficients of such a model are likely to be biased. To try to separate out effects, it is therefore first and foremost critically important to begin with an as well-specified model as possible, making use of relevant available explanatory variables. [7]

Dugundji and Walker [8] illustrate issues in the empirical estimation of a discrete choice model with network interdependencies using mixed generalized extreme value model structures with pseudo-panel data. Several modeling strategies are presented to highlight hypothesized interaction effects. In absence of true panel data on interaction between identifiable decision-makers, they use a priori beliefs about the social and spatial dimension of interactions to formulate the connectivity of the network and use socioeconomic data for each respondent as well as the geographic location of each respondent's residence to define aggregate interactions by grouping agents into geographic neighborhoods and into socioeconomic groups where the influence is assumed to be more likely. Technically, however, interactions between identifiable decision-makers may also be modeled using the approach described given the availability of suitable data.

In our empirical case study on adoption of Twitter clients, we do indeed have available data on which identifiable agents (Twitter users) plausibly influence other identifiable agents' choices, and furthermore we have longitudinal panel data observing repeated choices by agents over time. In this paper with such rich data, we continue this exploration of issues in the empirical estimation of discrete choice models with social interactions. Since our data is fairly large -more than 10,000 agents- we argue that the effect of unobserved correlated effects as perceived by any given agent is normally distributed, but is the same for that agent over the fairly short time period of the data collection. This simplified assumption allows us to specifically control for correlations in the error structure, through the use of mixed multinomial logit models with panel effects. [9].

2.4 Capturing unobserved correlated effects

Suppose each agent n makes a sequence of choices at a number of points in time indexed $(1,\ldots,t,\ldots,T_n)$. For our case study, we will consider a general case where the number T_n of decision-making moments per agent varies across agents. We introduce an additive, normally-distributed agent-specific error term for each alternative i as follows:

$$U_{int} = V_{int} + \varepsilon_{int} + \sigma_i \xi_{in} \ ; \ \boldsymbol{\xi}_n \sim N(0, I)$$

Conditional on $\boldsymbol{\xi}_n$, the probability that agent n makes a particular sequence of choices over time $(i_1,\ldots,i_{Tn})$ is given by the product of the probabilities for agent n making each individual choice i_t:

$$P_n(i_1,\ldots,i_{T_n} \mid \boldsymbol{\xi}_n) = \prod_{\forall t \in T_n} \frac{e^{\mu\left(V_{int} + \sigma_i \xi_{in}\right)}}{\sum_{\forall j \in C_n} e^{\mu\left(V_{jnt} + \sigma_j \xi_{jn}\right)}}$$

The unconditional user choice probability is the integral of this product over all values of $\boldsymbol{\xi}_n$

$$P_n(i_1,\ldots,i_{T_n}) = \int_{\boldsymbol{\xi}_n} \prod_{\forall t \in T_n} \frac{e^{\mu\left(V_{int} + \sigma_i \xi_{in}\right)}}{\sum_{\forall j \in C_n} e^{\mu\left(V_{jnt} + \sigma_j \xi_{jn}\right)}} N(0, I) d\boldsymbol{\xi}_n$$

2.5 Econometric estimation with simulation

The unconditional choice probability is approximated through simulation for any given value of ξ_{in} as follows:

1) Draw a vector of values of $\boldsymbol{\xi}_n$ from $N(0,I)$ for each alternative in the choice set C_n, and label this $\boldsymbol{\xi}_n^{\,r}$ with the superscript $r = 1$ referring to the first draw

2) Calculate the conditional user choice probability for the particular sequence of choices made by agent n with this draw

3) Repeat steps 1 & 2 for R total number of draws and average the results

$$\hat{P}_n(i_1,\ldots,i_{T_n}) = \frac{1}{R} \sum_{r=1}^{R} \prod_{\forall t \in T_n} \frac{e^{\mu\left(V_{int} + \sigma_i \xi_{in}^{\ r}\right)}}{\sum_{\forall j \in C_n} e^{\mu\left(V_{jnt} + \sigma_j \xi_{jn}^{\ r}\right)}}$$

If the estimated coefficients σ_i can be shown to be statistically insignificant, we assume that the hypothesized endogeneity has negligible effect.

3 Modeling the effects

This paper studies the diffusion of Twitter clients within the Indie community. Based on earlier research [10] we were able to determine a community of Indie developers that are actively using Twitter, using a mixed method community detection approach. For this community we use Twitter's publicly available API to gather data on network connections and actual messages sent. For 39 days, from 9 August until 16 September 2009, we harvested tweets and network connections on a daily basis for each of the nodes in the community.

Based on a review of the case study and the data [11], we expect client choice to be influenced by a number of distinct dimensions. Generally, social networks or social capital are considered to be important variables in explaining the adoption and diffusion of behavior. However, it is debatable to what extent the actual social connections, the global cultural discourse, and individual preferences influence this adoption and diffusion. Through our modeling of the effects, we can try to test the different hypotheses.

We distinguish between four Indie clients (Tweetie, Twitterific, Birdfeed and Twittelator), two popular non-Indie clients (Twitterfon and Tweetdeck) and the default Twitter web interface ("Web"). In addition, we employ a choice alternative, "Other" that serves as a baseline reference for the modeling. The "Other" category is highly heterogeneous and consists of more than 3500 clients that have relatively small market share (< 1%). On the basis of data we proceed to construct the following nine sets of different kinds of social and individual explanatory variables to explore in our model.

3.1 Contextual effects: taste maker influence

We start with exploring the contextual effect of whether or not a user in the community is connected to professional independent tech blogger John Gruber. Since Gruber promotes different clients to different extents [11], we are interested to see if the clients he promotes most favorably are used more often by the users connected to him. We operationalize this dummy variable in two different ways: if a user "follows" Gruber (ie. user receives tweets from Gruber); and if there is a reciprocal link with Gruber.

3.2 Contextual effects: developer influence

Next, we are interested in the contextual effect of whether or not a user in the community is connected to a Twitter client developer [11] as follows: Clients developed by "Indies": Tweetie (Loren Brichter, Atebits); Twitterrific (Craig Hockenberry, Iconfactory); Twittelator (Andrew Stone, Stone Design); Birdfeed (Buzz Andersen, SciFi HiFi); Clients developed by others: TweetDeck (Iain Dodsworth, TweetDeck); TwitterFon (Kazuho Okui, Naan Studio). We operationalize each of these dummy variables in two different ways: if a user "follows" the developer (ie. user receives tweets from the developer); and if there is a reciprocal link with the developer (ie. the link with the developer is especially strong).

3.3 Behavioral characteristics: power users

Since the Twitter clients have very different features, we might expect users who tweet a lot to prefer different kinds of clients than users who tweet less frequently. We operationalize this variable in four different ways: number of tweets sent by a user during observation period; "status count" (total tweets sent by a user during their entire history); number of tweets sent by a user prior to observation period (ie. giving emphasis of how active the user was in the past and how long the user has been using Twitter); and finally, the ratio of tweets sent by a user during observation period to total tweets sent during their entire history.

3.4 Network measures: central users

As per our review of the importance of social media networks for "echo-chamber" marketing, we are interested in whether a user's position in the community affects their client choice. We compute five classic network centrality measures: in-degree centrality (the number of a user's "friends" in sample, ie. from whom tweets are received); out-degree centrality (the number of a user's "followers" in sample, ie. to whom tweets are sent); closeness centrality (sum of distances from a user to all other users, giving an indication of the expected time until arrival for information that might be flowing through the network); betweenness centrality (how often a user lies along the shortest path between two other users, giving an indication of access to diversity of information); and finally, eigenvector centrality (measures if a user is connected to many users who are themselves well connected, identifying users in centers of cliques).

3.5 Network measures: extended user in-degree

In order to test the relative importance of the exposure to information flowing through the wider Twitter universe outside of the Indie community, we explore three extra network measure variables: the total number of a user's "friends" in the entire Twitter universe, ie. from whom a given user in principle receives tweets; the number of users outside the community from whom a given user in principle receives tweets; and finally, the ratio of users inside sample from whom a given user receives tweets to their total "friends" in the Twitter universe.

3.6 Network measures: extended user out-degree

Similarly, in order to test the relative opportunity to influence other users in the wider Twitter universe outside of the Indie community, we explore three extra network measure variables: the total number of a user's "followers" in the entire Twitter universe, ie. to whom a given user in principle sends tweets; the number of users outside the community to whom a given user in principle sends tweets; and finally, the ratio of users inside sample to whom a given user sends tweets to their total "followers" in the Twitter universe.

3.7 Temporal effects: individual preferences

We operationalize individual preference by constructing an alternative-specific relative individual cumulative lag variable. For each tweet, we count how often the sending user has been using each client in the seven days prior to sending the tweet resulting in an absolute cumulative lag variable. For each client, we then convert this absolute frequency to a relative cumulative lag variable indicating that client's use relative to how often that user has been using other Twitter clients in the past seven days. This individual preference variable shows how "sticky" a particular client has been for a user in the past seven days. This individual past behavior is likely to be a predictor of client choice for the next tweet, capturing complex UI preferences which we as researchers were not able to measure directly.

3.8 Temporal effects: social network contagion

To operationalize network influence we use the absolute cumulative lag variable as a basis. For each tweet, we count how often all users that the sender of that specific tweet is following use each client in the seven days prior to sending that the tweet. We convert the absolute frequency to an alternative specific relative network influence variable that indicates how often each client has been used relative to all other clients by all users that the sender of the tweet is following (ie. receiving information from). This can entail specific mentions of a client in a tweet but also more implicit or tacit knowledge about which client is popular or deemed useful within that user's social network. We argue that this usage by "friends" might influence client choice by either specific mentions of a client in Tweets or by the effect of tacit knowledge encoded within a user's social network.

3.9 Global influence

The cultural discourse on what is popular within the entire Indie community is operationalized by a set of alternative specific constants (ASC). Amongst things such as price and the impact of media exposure, we argue that this effectively captures global influence. It indicates the popularity of an alternative relative to all other alternatives during the entire sample period, after controlling for all other effects.

4 Results

All models are estimated using the freely available optimization toolkit Biogeme (http://biogeme.epfl.ch) developed by Bierlaire. We begin by estimating a baseline multinomial logit model with alternative specific constants only, representing global bias. The log likelihood, number of estimated parameters and adjusted rho-squared are given in the first line of Table 1.

Table 1: Log likelihood tests for incremental model specifications

Nr	Log Likelihood	Est. Par	Rho Sq	$-2[L_R-L_U]$	$\chi^2(0.1)$	p-Value
1	-968350.6	7	0.262	- -	- -	- -
2	-954368.7	14	0.272	27964	18.5	0.000
3	-568721.3	21	0.566	771295	18.5	0.000
4	-567945.3	28	0.567	1552	18.5	0.000
5	-566798.7	34	0.568	2293	16.8	0.000
6	-562010.9	41	0.571	9576	18.5	0.000
7	-561154.7	48	0.572	1712	18.5	0.000
8	-560664.6	55	0.572	980	18.5	0.000
9	-559662.6	62	0.573	2004	18.5	0.000
10	-559546.0	69	0.573	233	18.5	0.000
11	-452048.1	76	0.655	214996	18.5	0.000

1: Baseline model with alternative-specific constants only; 2: + Social network contagion (sq root); 3: + Lagged individual preferences (sq root); 4: + Follows Gruber; 5: + Follows developer; 6: + Frequency tweets during observation period (sq root); 7: + Eigenvector centrality (sq root); 8: + Closeness centrality; 9: + Ratio in-degree to total friends in Twitter (sq root); 10: + Ratio out-degree to total followers in Twitter (sq root); 11: + Estimated user-specific error component

Next we test one-by-one each of the explanatory variables defined in Section 3.1-3.8. In cases where the variables are continuous (ie. for all cases except for the dummy variables in section 3.1 and 3.2), we also test linear, quadratic and square root forms of these variables. Based on log likelihood tests compared to the baseline model and t-tests on the estimated coefficients [2], we identify the best fitting variables per category. For example, the dummies defined as "follows Gruber" and "follows developer" are more significant than their respective forms "reciprocal link with Gruber" and "reciprocal link with developer"; the most significant centrality measures are closeness and square root of eigenvector centrality, etc. The interested reader is referred to [11] for details and interpretation.

Having determined the best fitting variables and their respective functional forms, we then add the variables incrementally to the model, testing the improvement in log likelihood at each step. This is important to do, since variables that may have been significant when included in the model specification on their own, might no longer be significant when included together due to significance being shared between variables. The results are reported in lines 2-10 of Table 1. Each successive specification adds seven new parameters to the model (with the exception of "follows developer" where there are six since the Web alternative does not have a third party developer), as our data is rich and extensive enough to support alternative-specific definitions of the variables. In our case study, each new set of variables significantly improves the log likelihood (p-value of 0.000).

Finally, we include the normally-distributed user-specific error terms as in Section 2.4. We test the robustness of results using three different optimization algorithms for the maximization of the log likelihood, each with ten different random seeds for generating the draws. We use the estimated coefficients from the model in line 10 of Table 1 as a starting point for these 30 estimation runs with 50 draws, and then use the results with 50 draws in turn as the starting point for another 30 estimation runs with 200 draws, etc., for increasing number of draws, until the results stabilize across the random seeds for the three different optimization algorithms. Accounting for the unobserved correlated effects gave a dramatic jump in log likelihood as seen in line 11 of Table 1. The estimated coefficients in the final model in line 11 were also significantly different beyond standard error margins for 63 of 69 variables in the model in line 10. [11] Failing to account for unobserved correlated effects can thus yield misleading market share predictions for users' preferences for Twitter clients.

7 Conclusions and recommendations

A prominent approach to studying the dynamics of networks and behavior stems from a growing stream of research on stochastic actor-based models. See Snijders, van de Bunt, and Steglich [12] for a tutorial. With the large data in our case study however, these established methods are not tractable. The alternative approach we discuss in this paper allows us to apply other freely available, open source, existing software for the estimation of the models. In so doing, we hope to stimulate researchers and practitioners to adopt these techniques when using large data sets of more than 1000 nodes due to the relatively lower entry barrier than could be the case if dedicated code would need to be written or if expensive software would need to be purchased. An interesting direction for further discrete choice research on diffusion in large networks may be combining the approach of Aral, Muchnik and Sundararajan [13] for distinguishing causal effects using propensity score matched sample estimation in dynamic networked settings, with the present work accounting for unobserved correlated effects.

References

[1] M. van Meeteren (2008). Indie fever: The genesis, culture and economy of a community of independent software developers on the Macintosh OS X platform. A Sofa Publication [On-line]. Available: http://www.madebysofa.com/indiefever

[2] M. Ben-Akiva and S. R. Lerman (1985) *Discrete Choice Analysis: Theory and Application to Travel Demand*. Cambridge, MA: MIT Press.

[3] D. McFadden (2001) Economic choices. *American Economic Review* 91(3): 351-378.

[4] D. McFadden (2010) Sociality, rationality, and the ecology of choice. In S. Hess and A. Daly (eds.), *Choice Modelling: The State-of-the-Art and the State-of-Practice*. Bingley, UK: Emerald Group Publ. Ltd.

[5] W. A. Brock and S. N. Durlauf (2002) A multinomial choice model of neighborhood effects. *American Economic Review* 92(2): 298-303.

[6] C. F. Manski (1995) *Identification Problems in the Social Sciences*. Cambridge, MA: Harvard Univ. Press.

[7] E. R. Dugundji (2012) *Socio-Dynamic Discrete Choice: Theory and Application*. Ph.D. manuscript. Universiteit van Amsterdam, Netherlands.

[8] E. R. Dugundji and J. L. Walker (2005) Discrete choice with social and spatial network interdependencies: An empirical example using mixed generalized extreme value models with field and panel effects. *Transportation Research Record* 1921:70-78.

[9] K. E. Train (2009) *Discrete Choice Methods with Simulation, 2nd ed*. New York: Cambridge University Press.

[10] M. Meeteren, A. Poorthuis, and E. R. Dugundji (2010) Mapping communities in large virtual social networks, IEEE International Workshop on Business Applications of Social Network Analysis, Bangalore, India. [Online]. Available: http://dx.doi.org/10.1109/BASNA.2010.5730297.

[11] E. R. Dugundji, A. Poorthuis, and M. van Meeteren (2012) Capturing unobserved correlated effects in diffusion in large virtual networks: Distinguishing individual preferences, social connections and cultural discourse influence on the adoption of Twitter clients, unpublished.

[12] T. A. B. Snijders, G. G. van de Bunt, and C. E. G. Steglich (2010) Introduction to stochastic actor-based models for network dynamics. *Social Networks* 32(1): 44-60.

[13] S. Aral, L. Muchnik and A. Sundararajan (2009) Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proc. Nat'l Acad. Sci.* 106(51): 21544-21549.