Dynamic Ego Networks

David Hachen,[1*] Omar Lizardo,[1] Others[2]

[1]Department of Sociology, University of Notre Dame,
810 Flanner Hall, Notre Dame, IN 46556
[2]Another Unknown Address, Palookaville, ST 99999, USA

[*]corresponding author: David Hachen, dhachen@nd.edu.

1 Introduction

Networks including social networks consist of vertices (nodes) and edges connecting vertices. One of the most important if not the most important property of vertices is their degree, also referred to a degree centrality, ego network size, network range, and popularity[1]. Vertex degree captures the the number of immediate network neighbors (alters) that a vertex (ego) has, that is those nodes that are "adjacent" to an ego and which, therefore, a person can directly interact. At the level of the network, one of the most important global properties of networks is the degree distribution which is rarely uniform. There is variance (inequality) in degree with some persons having many alters and others very few. While there are numerous expectations

[1]Popularity is usually defined in terms of a person's in-degree, i.e. the number of alters that consider the ego to be their friend.

concerning the shape of degree distributions, and theories about why certain shapes are more probable and how they emerge [REF][2], no one disagrees with the proposition that degree distributions evidence inequality.

Substantively, past research has shown that in social networks, vertex degree is an important status marker that predicts a variety of usually beneficial individual level outcomes [REF] and is correlated with other measures of status such as socio-economic status [REF]. In organizational settings, ego network size is indicative of the centrality of a person, that is the extent to which the person "holds together" a set of interacting persons and plays a role as an intermediary [Burt, REF][3]. In diffusion studies, the importance of hubs (high degree vertices) for the spreading of some trait throughout a network (and the speed at which diffusion occurs) is well documented [REF]. Network size also tends to be correlated with other important ego network traits. It is usually positively correlated with ego network diversity (defined in terms of various node traits such as demographic traits, attitudes and tastes) and negatively correlated with density (i.e. an ego network's clustering coefficient). [REF]

Though there are theories, models and simulations about how the degree of a vertex changes over time [REF], empirical over time analyses are rare because of the paucity of data that allows researchers to track changes in the degree of persons within a population. In order to understand why and

[2]go over PA models, power laws, gaussian distribution, etc

[3]note on various types of centrality and high corelation with degree, ref to dynamic centrality measures

how people vary in the size of their networks at one point in time we need to look at how their networks emerge over time and the factors that affect and shape ego network size growth patterns. The ideal "natural experiment" would be to put individuals who have no prior ties together in some place and observe social interactions in order to measure when ties form and dissolve and compute changes over time in the size of networks. We expect that such an ego network degree time-series would evidence an initial growth period in which tie formation rates are greater than tie dissolution rates. It is the period after this "growth period" (whose length could vary across individuals) that is of interest. With data on tie formation and dissolution that allows us to date the beginning and end of a tie's social "life", we could address two important questions pertaining to ego network dynamics:

- How stable is the size of a person's ego network after the initial growth period?
- How much variance is there across person's in the size of their networks?

The first question pertains to the within subject across time variability in network size, the latter question to the between subject variability. So the overall question is how much the total variation across time and people is due to within subject over time variability and how much is the result of between subject variability.[4]

[4]in addition to investigating within and betwen subject variability, we could also map out and try to predict variation in qualities of the growth period, the period when the growth rate is increasing, such as its duration and the shape of the growth curve during the growth period.

Answering these questions requires that we take time into account when measuring the degree of a person at any particular point in time. All measures of degree, including "static" measures of degree that are typically used in cross-sectional research, make assumptions about temporal processes, but rarely are these assumptions made explicit. In Section 2 we make these assumptions explicit by distinguishing between three different measures of degree: cumulative degree, effective degree, and active degree. Based on the work of Miritello, we show how these different measures of degree take into account tie formation and dissolution in computing degree [Burt cite???], how temporally fine-grained dynamic network data are used to compute these measures, and how these degree indicators are dependent on the temporal window within which individuals are observed. We also show why effective degree (what Mireitello et. al. call the instantaneous degree) is the critical quantity because it captures the number of people an ego *could* interact with at an given moment in time. We also show why cumulative degree, the total number of people a person *has* interacted with during some time period, is a problematic measure of degree because it conflates two important quantities: effective degree and churn (i.e. turnover in the people who are in an ego's immediate (1-hop,adjacent) social network).[5]

[5] In cross-sectional research, cumulative degree is determined in one of two ways. When using name generators in surveys where people are asked to self-report the names of all the people they interacted with (sought advice from, talked aobut important matters with) over some period of time (e.g., the last two months). Degree then is just the count of the number of people a person named, i.e. the total number of people that a person interacted with (as specified by the name generator) during some time period. If behavioral data are used (such as observation data on between subject interactions or traces of such

In Section 3 we discuss the data we use to examine the within and between subject variance in effective degree, and how our data was generated as part of the NetSense Study. This study was designed to come as close as possible to the natural experiment setup in which subjects without any prior connections come together and form and dissolve ties over time. In Section 3 we discuss how we selected a cohort of approximately 200 first-year college students at the University of Notre Dame and obtained data on the communicative interactions they had with other people, including other ND students. These communication records allow us to map the evolution of each person's social networks over a 70 week period. In Section 4 we address our two research questions by presenting results concerning network size evolution, temporal stability and between-subject variation. We show that the three measures of degree have different temporal signatures and patterns indicative of period effects, and we assess the extent to which effective degree varies across time and/or across persons. Section 5 concludes by discussing the implications of this dynamic perspective on degree, how different measures of degree size capture different aspects of people's social networks, and future research directions that will allow us to ascertain the role that endogenous network process and/or exogenous factors (such as vertex traits) play in ego network size evolution and in predicting between-subject variation.

interactions such as logs of communication activity), degree is just the cumulative count of the number of people a person interacted with from the beginning to the end of the observation (or recording) period.

2 Measuring Degree

In order to measure degree dynamically we must incorporate time into the analysis. Each individual is observed on a sequential number of time units up until the last observation point $\{t_0, t_1, t_2, t_3 ... t_\omega\}$. These time units may be hours, days, weeks, months, etc. The time window in which we observe tie formation and dissolution is defined as $\Omega = \{t_0, t_1, t_2, ... t_\omega\}$. In this dynamic context we can distinguish at least three empirically and conceptually distinct notions of vertex degree:

- **Cumulative Degree** $k_i(T_k)$: Count of the number of *unique contacts* that a given individual has interacted with up until a given time point T_k.

- **Effective Degree** $\kappa_i(t_k)$: The number of *active contacts* that an individual has the *potential* to interact with at a given time point t_k.

- **Active Degree** $\kappa_i^*(t_k)$: The number of *active contacts* than an individual *does in fact interact with* at a given time point t_k.

In order to use the time window Ω to compute each of these measures we need to "date" ties, that is identify when they begin (the first interaction) and when they end (the last interaction). However because of left-censoring, we can not assume that the first interaction between an ego (i) and an alter (j) at time $t_k \in \Omega$ is indicative of a new tie. It could just be the activation of an old tie. Similarly, because of right-censoring, we can not assume that the last interaction at time $t_k \in \Omega$ is indicative of tie dissolution. The tie could just not be active for the remainder of the Ω time window. Miritello et al. (2013) solve this problem by using additional time windows before and after Ω to establish whether the first interaction in Ω indicates a new tie and the last one a deactivation.[6]

For left censoring, begin by identifying all interactions an ego has with all her alters within $\Omega = \{t_0, t_1, t_2, ...t_\omega\}$. Refer to these interactions by a tie variable $a_{ij}(t_k)$ which takes on the value of 1 if there is an interaction between i and j during t_k, otherwise 0. For each identified interaction, determine whether ego had an interaction with that alter in the time period prior to Ω (i.e. $a_{ij}(t_k) = 1$ for a value of t_k prior to Ω). Count the number of alters that ego had both an intereaction with within Ω and prior to Ω. This count is the number of persons an ego had ties with prior to the beginning of Ω, $\kappa_i(t_0)$. For all other interactions with an alter within Ω, identify the time point $t_a \in \Omega$ when the interaction first occurs. This is the *formation* date of the tie.

For right censoring, identify all ties where there is at least one interaction within Ω. Then determine if another interaction occurs in the specified time period after Ω (i.e. $a_{ij}(t_k) = 1$ for a value of t_k after Ω). If an interaction

[6] These additional time windows need to be long enough so that the probability of a tie not appearing in these time periods but having already been established (in the case of left-censoring) or will be active later (in the case of right-censorigG) is very low. Miritello et al. (2013) do this by computing the inter-event ties for all interactions and creating time periods before and after Ω that are long enough to encompase most of these inter-event times.

does occur after Ω, that tie is a *persistent* tie (does not decay within Ω).If not, identify the latest time point $t_w \in \Omega$ when tie was active. This is the *deactivation date* of the tie.

With information for each node on $\kappa_i(t_0)$ (the number of alters that ego interacts with within Ω and prior to Ω), the beginning date for all other ties, and the ending date for all ties that do not persist after Ω, we are able to compute the three degree measures.

2.1 Effective Degree

For *effective Degree* (κ_i), begin at t_0, where if there is left-censoring $\kappa_i(t_0) > 0$. For each discrete time interval within Ω (hours, days, weeks, months) compute the number of newly formed ties, $n_{\alpha,i}(t_a)$, in each discrete time period and the number of deactivated ties $n_{\omega,i}(t_w)$. Then the effective size for each discrete time period (e.g., week) is computed by adding to $\kappa_i(t_0)$ the number of new ties an ego had in period 1 ($n_{\alpha,i}(t_1)$ and substracting out the number of deactivated ties ($n_{\omega,i}(t_1)$). In general the effective degree for the ith person in discrete time period t_k is:

$$\kappa_i(t_k) = \kappa_i(t_{k-1}) + [n_{\alpha,i}(t_k) - n_{\omega,i}(t_k)] \quad (1)$$

It is evident from equation 1 that changes in $\kappa_i(t_k)$ are solely a function of the difference between the number of ties formed $n_{\alpha,i}(t_k)$ and number of ties deactivated $n_{\omega,i}(t_k)$. We refer to this difference as the growth rate $G_i(t_k)$:

$$G_i(t_k) = n_{\alpha,i}(t_k) - n_{\omega,i}(t_k) \quad (2)$$

which implies that:

$$\kappa_i(t_k) = \kappa_i(t_{k-1}) + G_i(t_k) \tag{3}$$

Therefore, for any given time period Δ that begins at t_a and ends at t_w:

1. Whenever $\sum_{t_a}^{t_w} n_{\alpha,i}(t) \neq \sum_{t_a}^{t_w} n_{\omega,i}(t)$, we should observe fluctuations in κ_i within persons over time.

2. Whenever $\sum_{t_a}^{t_w} n_{\alpha,i}(t) \simeq \sum_{t_a}^{t_w} n_{\omega,i}(t)$ we should observe little variation in κ_i within persons over time.

3. If (2) holds, then $\kappa_i(t_k) \simeq \bar{\kappa_i}$, where $\bar{\kappa_i} = \frac{1}{T}\sum_{t_a}^{t_w} \kappa_i(t)$ and T is the number of time points in Δ.

In sum, effective degree ($\kappa_i(t_k)$) measures the size of the pool of people with whom an ego *can* interact at t_k. This set of potential interactorss is specified by identifying all alters with whom ego has interacted with prior to t_k *and* will interact with during or after t_k. The size of this set, $\kappa_i(t_k)$, can vary either over time and/or across persons. Effective network size will vary over time (i.e. grow or decline) if the number of new ties formed is not the same as the number of ties that are dissolved.

2.2 Cumulative Degree

The cumulative degree, $k_i(T_k)$, is the total number of alters a person has interacted with up until time k. Set $a^*_{ij}(t_k) = 1$ when an interaction between i and j occurs for the *first time* in Ω. Then for a given discrete time period, t_k, $k_i(t_k) = \sum_j a^*_{ij}(t_k)$. This is just the count of the number of interactions an ego has with *different* alters *for the first time* in t_k. The cumulative degree is then computed as:

$$k_i(T_w) = \sum_{t_0}^{t_w} k_i(t_k) \tag{4}$$

Note that the cumulative degree differs from the effective degree in two important respects. First, it ignores left-censoring so that the first time an interaction occurs with an alter is counted as the first interaction. As a result $K_i(T_0) = 0$. Second, tie dissolution is ignored. What is cumulated is the count of the number of first time incidences of interactions (i.e. "new" ties, some of which could have been active prior to Ω). This means that another way to express the cumulative degree is:

$$k_i(T_w) = \kappa_i(0) + \sum_{t_0}^{t_w} n_{\alpha,i}(t_k) \tag{5}$$

That is, the cumulative degree is the sum of the number of ties that an ego had prior to Ω (and which were active within Ω) and the total number of new ties formed within Ω.

Given that the calculation of a person's effective degree involves subtracting out deactivated ties, a person's cumulative degree must be equal to or larger than their effective degree. If a person did not form any new ties or dissolve any ties that existed prior to Ω (i.e. $n_{\alpha,i}(t_k) = n_{\omega,i}(t_k) = 0$), $\kappa_i(t_k)$ would not vary over time (i.e. $\kappa_i(t_k) = \bar{\kappa_i}$) and would be equal to the value of $\kappa_i(t_0)$ (i.e, $k_i(T_k) = \bar{\kappa_i} = \kappa_i(t_0)$). If however new ties are formed and old ties deactivated, $\kappa_i(t_k) < k_i(T_k)$ because the cumulative degree counts prior ties that are active in Ω ($\kappa_i(t_0)$) and new ties ($\sum_{t_0}^{t_w} n_{\alpha,i}(t_k)$), but ignore deactivated ties. As we discuss below, when there is churn within ego networks as new ties replace deactivated ties, the cumulative degree is a function both of effective degree and the level of churn.

2.3 Active Degree

Active degree, κ_i^*, is the total number of alters an ego interacts with in a discrete time period, t_k. To compute active degree we use the indicator tie variable described earlier, $a_{ij}(t_k)$, which takes on the value of 1 if i interacts with j at t_k, otherwise $a_{ij}(t_k) = 0$. Then the active degree is just the sum of these tie variables over all of i's alters j for a discrete time period:

$$\kappa_i^*(t_k) = \sum_j a_{ij}(t_k) \tag{6}$$

$\kappa_i^*(t_k)$ captures how frequently an ego interacts with those who are in her pool of alters as defined by $kappa_i(t_k)$. If a person interacts in the time period t_k with all of the people she could interact with, then $\kappa_i^*(t_k) = kappa_i(t_k)$. As such we can normalize active degree by dividing it by $\kappa_i(t_k)$:

$$\kappa_i^{**}(t_k) = \frac{\kappa_i^*(t_k)}{\kappa_i(t_k)} \tag{7}$$

This normalized active degree captures the proportion of a person's effective degree network that she interacts with in a given discrete time period t_k. The concept of active degree is related to the strength of a social tie. Persons with stronger ties to their alters will have higher average active degrees than those who have weaker ties. Variation over time in the normalized active degree are indicate of period effects on activity level, while variation across persons in

$kappa_i^{**}(t_k)$ indicate the extent to which people differ in the average strength of their ties to their alters.

2.4 Churn

As noted earlier, cumulative degree ($k_i(T_k)$) is likely to overstate effective degree ($kappa_i(t_k)$) because cumulative degree is a function of both a person's effective degree and churn in their ego network. Churn captures the extent of change (turnover) in who is in the pool of people with whom an ego could interact. If in a given time period there is no formation of new ties or dissolution of existing ties, then $n_{\alpha,i}(t_k) = n_{\omega,i}(t_k) = 0$ and $\kappa_i(t_k) = \kappa_i(t_0) = \bar{\kappa_i} = k_i(T_k)$. When new ties replace deactivated ties so that $n_{\alpha,i}(t_k) = n_{\omega,i}(t_k)$[7] then $\kappa_i(t_k)$ will be stable ($\kappa_i(t_k) = \kappa_i(t_0) = \bar{\kappa_i}$) and churn will just be the count of the number of new ties (which is the same as the number of deactivated ties):

$$C_i(T_w) = \sum_{t_0}^{t_w} n_{\alpha,i}(t) = \sum_{t_0}^{t_w} n_{\omega,i}(t) \tag{8}$$

In situations where there is growth in a person's ego network ($n_{\alpha,i}(t_k) > n_{\omega,i}(t_k)$) or decline ($n_{\alpha,i}(t_k) < n_{\omega,i}(t_k)$) computing churn is more complicated. In the case of grow some of the newly formed ties are indicative of growth while others are replacing deactivated ties. As such churn is the count of deactivated ties (which are replaced with new ties): $C_i(T_w) = \sum_{t_0}^{t_w} n_{\omega,i}(t)$. In the case of decline, some deactivated ties are indicative of decline while

[7]NOTE: need to do this over an extended period of time and cumulative new and decay ties. For a small time period it is unliekly that they will differ in value. Ie can be lags

others are replaced with new ties (i.e. churn). As such churn is the count of new ties: $C_i(T_w) = \sum_{t_0}^{t_w} n_{\alpha,i}(t)$.

When effective network size is fairly stable over time because the number of new ties is the same as the number of deactivated ties, then we can normalize churn by dividing it by a person's effective degree:

$$C_i^*(T_w) = \frac{C_i(T_w)}{\bar{\kappa}_i} \tag{9}$$

This ratio captures the proportion of a person's network that turned over in a given time period.

We can now see how cumulative degree is related to effective degree and churn. Recall from equatin 5 that the cumulative degree is equal to a person's effective degree at the beginning of Ω plus the number of new ties. As we have seen in equation 8 Churn, under conditions when $\kappa_i(t_k)$ is stable, is just the number of new ties. Therefore a person's cumulative degree can be expressed as:

$$k_i(T_w) = \kappa_i(0) + C_i(T_w) \tag{10}$$

A person's cumulative degree overstates their effective degree to the extent that there is churn in their social network.

3 The NetSense Data

To compute the various measures of degree we need longitudinal data so that we can identify the beginning and ending dates of all ties. The data used in this analysis comes from the NetSense study, a two-year data collection effort (now extended to 4 years) funded by the National Science Foundation and subsidized by Sprint Telecommunications. 196 first year students who began attending the University of Notre Dame in the Fall of 2011 were given a smartphone and a free voice/text/data plan. We created a monitoring agent that was installed on all phones which captured technical information (when the phone was on and off, whether it was using Wi-Fi or 3G) as well as information on all communicative interactions that occurred through the phone, including voice calls, SMS text messages, and email sent and received through the phone. The focus of this analysis is on the text messages and voice calls. For each text and phone call sent or received by one of our smartphones a call data record (CDR) is generated which includes the sender and receive phone numbers, the time and date when the communication began, the duration of the event (if a voice call), and the type of communication (voice or text). No information on the content of text message or phone calls was recorded. We view these CDRs as unobtrusive indicators of social interaction between one of our NetSense Study participants and one of their alters that can be used to construct their ego networks and to examine change in their ego networks over time. While there are many other ways in which a student can interact with someone else (via email, through social media such as Facebook and Twitter, and through face-to-face meetings), given the high prevelance of texting and the correlation of texting and meeting up with friends, we believe that measures of network interactions and connections derived from calls and text to be very reliable and valid.[8]

In addition to the CDR data we obtained data from our participants through two types of surveys: (1) a standard survey through which we gathered information on each student's views, aspirations, attitudes, tastes, and behaviors during a specific time period, (2) a network survey in which we used a broad name generator to elicit the names of up to 20 people that they had interacted with in prior months, along with their phone numbers and basic information on each of their contacts. Both of these types of surveys were administered prior to students arriving on Campus in the Fall of 2011 and then every 3-4 months thereafter. The sampling design through which we obtained the 196 subjects was designed so that we would have a good number of first year students from a small subset of residence halls. First year ND students are randomly allocated to one of 29 same-sex dorms. We initially chose 4 dorms (2 male and 2 female) that were located near each other and sent via email invitations to participate in the NetSense study to all first year students in these dorms. We then opened the study up to first year students in 4 other dorms located near the original dorms, and then finally to all first year students in any of the dorms. The result is that we have students from each of the 29 dorms, but a few dorms are overrepresented. Four residence halls have 15 or more NetSense subjects, while 7 dorms has 6-9 NetSense

[8] In other work our team investigates the correlation between social network adjaceny matrices derived from emails, phone calls, text messages, Facebook posts, and face-to-face interaction as indicated by proximity of two phones using Bluetooth detection technologies. See REF

subjects, and the remaining 18 dorms have 1-5 students. Comparison of our sample to the full population of first year ND students indicates that our sample is fairly representative of the 2011 incoming class[9]. Our NetSense participants tend to come from slightly less well-off backgrounds in terms of both education and income, and are more likely to be African-Americans, Asian-Americans or Latinos. The lower SES levels and greater representation of non-white groups is probably due to the fact that students who did not already have a high-level smartphone were more likely to accept our invitation. Of the 196 students who started the study, 161 were still in the study at the end of the first academic year. For most of these students we have CDRs for 106 weeks. During this period these students were involved in approximately 6 million communication events with almost 55,000 alters. In our analysis we only focus on mutual dyads, that is dyads in which i initiated at least one communication with j and j initiated at least one communication with i. About 25,000 dyads are mutual. The vast majority of these ties and communication events are with people who are not in the NetSense study.

We want to look at new tie formation after arrival at Notre Dame. To do this we restrict our analysis to ties our NetSense subjects have with other ND students. We are able to identify whether one of our subject's alters is a ND student using the network survey data. In the network surveys, students are asked to list the people they have interacted with, give us their phone

[9]We compare our sample's distribution on a few key variables to the Cooperative Institutional Research Program data for Notre Dame which is derived from surveys administered to all incoming students prior to arrival.

number, and tell us whether the person is another ND student. Using the second and third network surveys administered at the end of the Fall 2011 and Spring 2012 semester, almost 13,000 ties (46% of all mutual ties) are identified as ties to ND students.[10] Using information obtained from the network surveys, we are able to identify two other type of ties, both of which existed prior to arrival on campus: family ties which are about 3% of all mutual ties, and non-family ties to high school friends and others that our subjects said they knew prior to arrival at ND which are about 5% of the mutual ties. For the remaining ties (approximately 43% of the mutual ties) we know nothing about the alter. In the analysis we use only on information on the post-ND ties to construct the size of each student's network. By restricting the analysis to ties with alters who are also ND students we solve the left-censoring problem given that all of these ties began after arrival on campus. By looking at the evolution of each NetSense participants ND ego network we are able to observe and measure how these ego networks evolve from a state in which each of our subjects has no ties.

The time period, Ω, that we use in this study covers the first 70 weeks of the study (the Academic Year 2011-12, Summer 2012, and Fall semester 2012). We use the subsequent 36 weeks (Winter break 2012, Spring 2013 semester and Summer 2013) to determine if the last week in which the tie

[10] We also identified ties to ND students through three other means: whether a phone number is in the CDRs of two unrelated NetSense subjects, whether the phone number is associated with an ND email address according to the address book on the phone of a NetSense participant, or whether the phone number appears in the address books of two unrelated NetSense study participants.

is active in Ω is a case of tie deactivation. If the tie is active in the first 70 weeks but not active in the subsequent 36 weeks we consider the tie to have been activated and deactivated in the first 70 weeks and we mark its last activation week as its ending date. Of the 12,565 post-arrival ND ties, 91% were activated during the first 70 weeks. The remainder were activated after that and are not considered in this analysis. Of these 11,401 ties, 62% were not active during weeks 71-106 and are, therefore, considered to have been deactivated during the first 70 weeks.

We carve up Ω into weeks [11] and use the CDR data to identify the first week in which one of our subjects interacted by voice or text with another ND student. The identified week is considered to be the starting week for that tie. Note that with this setup there is no left censoring. Incoming Notre Dame students are assumed to have had no interactions with other ND students prior to arrival on campus.[12] Therefore, the week in which the first interaction with a ND alter occurs is the beginning week for that tie.

[11] A week goes from Sunday to Sunday.

[12] Notre Dame is a national university drawing students from across the United States (and other countries). While a few of our subjects may have knows a few other ND students prior to arrival (e.g., family members, students from the same area, students they met online or when visiting campus) all indications are that such contacts are few in number.

4 Findings

4.1 Effective Degree

Figure 1 plots changes over the 70 week period in the means values of Cumulative Degree ($k_i(T_k)$), Cumulative Deactivated Ties, Effective Degree ($\kappa_i(t_k)$), and Active Degree (κ_i^*). All students begin with no ties, but by the end of the first week the mean number of new ties in our population is almost 4.[13] Means values for cumulative degree continue to rise throughout the 70 week period as students form new ties with other students. By the end of 70 weeks the typical student has formed about 64 ties with other students. Tie formation evidences period effects as students form very few new ties during the last 2 weeks of a semester (weeks 18-19, 39-40 and 70) and during breaks (weeks 20-23, 31-32, and 41-53). Tie decay (the red line in Figure 1 for Cumulative deactive count) begins fairly early, consistent with the liability of newness of ties [REF]. By week 5, the mean cumulative number of deactivated ties is almost 3. De-activations also evidence period effects. They tend to pick up at the end of each semester as students with whom an ego has interacted with during the semester do not continue interacting with them after a break period. By the end of 70 weeks, the typical student has deactivated about 38 of the 64 ties that they formed during this period (60%).

What is most interesting is the pattern for effective degree, ($\kappa_i(t_k)$). Af-

[13]There were a few ties that were created in the first week that also were never active after that, resulting in the $\bar{\kappa}_i(t_1)$ being less than $\bar{k}_i(T_1)$.

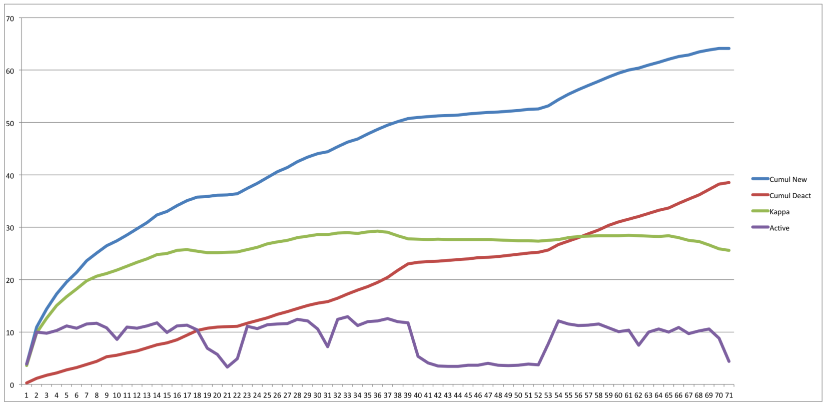

Figure 1: Mean weekly values of Cumulative Degree ($k_i(T_k)$), Cumulative Deactivated Ties, Effective Degree ($\kappa_i(t_k)$), and Active Degree (κ_i^*) over 70 weeks

ter rising during the first 15 weeks as students form many more new ties than they deactivate, $\kappa_i(t_k)$ stabilizes at around 25. While there is a small decline over winter break as students deactivate more ties than they form, effective degree increases at the beginning of the Spring 2012 semester and then stabilizes at around 28-29 until the very end of that semester when it drops slightly. Surprisingly after summer break (the beginning of the Fall 2012 semester) $\kappa_i(t_k)$ does not increase like it did at the beginning of the Spring 2011 semester. In fact by the end of the Fall 2012 semester effective degree drops to the levels seen in the Fall 2011 semester.

While there appears to be some change over time in mean effective degree, what is remarkable is how stable effective degree is, a finding consistent with the results reported by Miritello [REF]. After starting at zero, the typical student reaches their effective degree level by around week 15. From then on the typical student's degree remains fairly stable, with minor increases

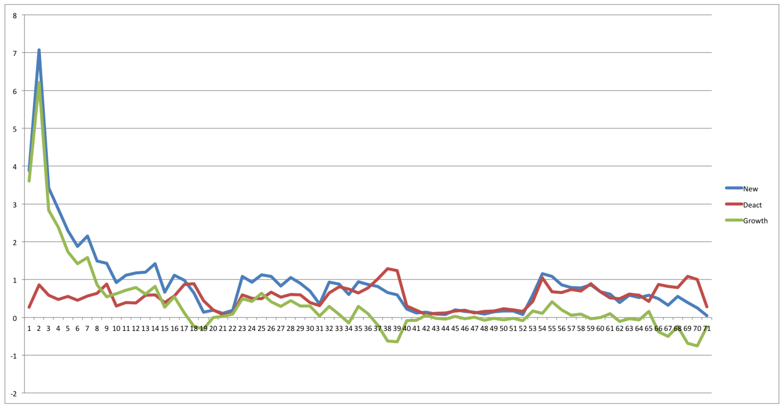

Figure 2: Mean Weekly Values of Number of New Ties ($n_{\alpha,i}(t_k)$), Deactivated Ties ($n_{\omega,i}(t_k)$), and Growth in Number of Ties ($G_i(t_k)$)

and decreases here and there. This implies that after an initial period of tie formation, new tie formation is counterbalanced by old ties being dissolved. While there are periods where tie formation rates differ from tie dissolution rates, overall what we observe are fairly stable effective degree levels.

Figure 2 provides insights into why effective degree is so stable. Recall from equation 3 that changes in effective degree are solely a function of growth, that is the difference between the number of new ties a person has in a given week and the number of ties that they deactivated. As seen in Figure 2 growth levels are very high in the first 5 weeks, with the typical student a net growth of 2-4 ties per week during this period. Deactivation levels, however, are not negligible. Though students continue to form new ties, they also dissolve other ties. Around week 18 dissolution outpaces formation, and we see negative growth (decline). During their second semester (Spring 2012), the typical student starts forming more new ties, and even though tie dissolution is still occurring, tie formation outpaces dissolution resulting in growth and increases in the typical student's effective degree. Again at the end of the semester we see an increase in tie decay leading to a decline in degree. Through the summer when there is little tie formation or dissolution effective degree remains constant, only to increase a bit at the beginning of the Fall 2012 semester before it declines at the end of the semester when dissolution rates outpace tie formation.

These trends indicate that there is a temporal dimension to tie formation and dissolution that creates fluctuations in effective degree. In a college setting, these period effects occur at the beginning and end of semesters when students "adjust" their networks. Furthermore, there appears to be two periods in which students grow their networks: upon arrival and then to a lesser extent during the beginning of the Spring 2012 semester, their second semester on campus. While some of the growth that occurs in Spring 2012 is counterbalanced by decline in people's networks at the end of the semester, there is enough growth so that effective network size is slightly larger at the end of the Spring 2012 semester than during the Fall 2011 semester. After the Spring semester there is virtually no growth (or decline) until the very end of the Fall 2012 semester when network size evidences a small decline. So while there are fluctuations over time in effective degree, what is most remarkable is how stable it is both within semesters (after the beginning and before the end of each semester), and across semesters.

The stability of effective degree is evident in the inter-temporal corre-

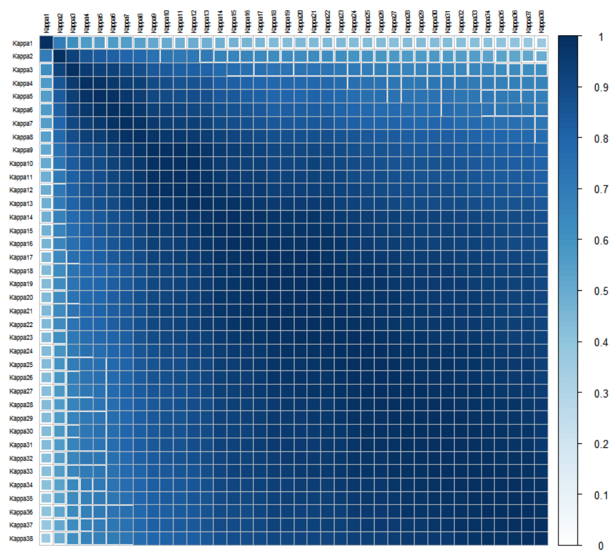

Figure 3: Inter-temporal Spearman Rank Correlations of Effective Degree (κ_i)

lations between values of $\kappa_i(t_k)$ for all pairs of weeks. Figure 3 displays these Spearman rank correlations [O - is this correct] for the first 38 weeks [NOTE:need to update this figure so that is covers 71 weeks] and shows that an effective degree value in any week except the first few weeks is highly correlated with the effective degree value for that person in all other weeks. This means that where a person stands in the effective degree distributin after a few weeks at school is an extremely good predictor of their rank in the effective degree distribution in future weeks, even weeks that are over 6 months away.

Because of this high stability in effective degree especially during semesters, we can ignore the temporal variation within semesters and compute the mean

			Correlations (Spearman)	
	Mean	St. Deviation	Weeks 26-36	Weeks 56-65
Weeks 13-18	25.09	13.84	0.94	0.74
Weeks 26-36	28.49	16	1	0.81
Weeks 56-65	28.35	18.7		

Table 1: Mean, Standard Deviation and Correlations Among Values of Effective Degree at Three Points in Time

values for each person of their effective degree during the semester. For the Fall 2011 semester we compute this mean over weeks 13-18 ($\bar{\kappa}_i(t_{13}, t_{18})$), the period after the rapid tie formation period (weeks 1-12) and after Fall Break. For the Spring 2012 semester we compute this mean over weeks 26-36, while for the Fall 2012 semester we use weeks 56-65. As seen in Table 1 the mean level of effective degree increases a bit from Fall 2011 to Spring 2011 and then remains the same in the Fall 2012 semester. The correlations indicate that a person's place in the degree distribution in the Fall 2011 predicts their place in that distribution in both the Spring 2012 and Fall 2012 semesters, though the correlation is lower for Fall 2011 to Fall 2012.

4.2 Cumulative Degree and Churn

Recall that the measure of cumulative degree captures new tie formation but ignores tie dissolution. As such it captures both effective degree and churn (see equation 5). This is seen in Figure 1 where after the first semester, the cumulative degree and cumulative deactivation curves are fairly parallel, while the line representing effective degree is fairly flat. While new ties are

			Correlations (Spearman)	
	Mean	St. Deviation	Weeks 26-36	Weeks 56-65
Weeks 13-18	0.244	0.156	0.24	-0.04
Weeks 26-36	0.328	0.166	1	0.034
Weeks 56-65	0.251	0.199		

Table 2: Mean, Standard Deviation and Correlations Among Values of Normalized Churn at Three Points in Time

formed and old ties dissolve, this is not changing the typical student's effective degree very much. Instead tie formation and dissolution are indicative of churn in people's networks. To see this Table 2 reports the normalized churn values for the weeks in each semester where effective degree is very stable. In Fall of 2011 the typical student experienced about a 25% turnover in their network, 33% in Spring 2012 and 25% in Fall 2012. While there is a very small positive correlation between Churn levels in the Fall and Spring of the first academic year, there is no association between levels in either of these semesters and the level in Fall 2012. Furthermore there is no association between a student's effective degree and level of churn in any of the semesters, a finding also reported by Miritello [REF]. While effective degree evidences a great deal of stability after the initial period in which students form their ties, churn levels vary over time so that levels of churn during one period do not predict churn levels later. Nevertheless, churn is a pervasive aspect of people's social networks. Though the size of ego networks may remain fairly stable over time, stability in the size of people's networks does not imply stability in who is in their ego network.

As such cumulative degree, the measure of degree that is often reported in social network studies, overstates a typical student's effective degree. By the end of 70 weeks, the typical student has formed ties with 64 other students, but has an effective degree of only 26, which means that they deactivated 38 ties during this period. [14] This means that the typical student with an effective degree of around 26 has a churn rate of around 150% in their network over this 70 week period. Because of this churn, the cumulative degree overstates effective degree.

4.3 Active Degree

As seen in Figure 1, active degree (κ_i^*), an indicator of how frequently a person interacts with their alters, evidences a good deal more temporal fluctuation than effective degree. During break periods, effective degree declines substantially, but picks up during semesters and remains fairly constant at around 11, with some slightly higher periods of activity right after a break. Active degree is less stable than effective degree as seen in Figure 4. Activity during the semester is not a good predictor of activity during breaks, indicating that during breaks some students may be more active in keeping in touch with ND friends than other students. Active degree is also lower in value that effective degree, indicating that the typical student is not able to keep in touch with all their ND friends every week. For example, in the Spring 2012 semester during weeks 26-36 the typical student interacted with

[14] In most applications where there is left censoring $\kappa_i(t_0) > 0$ and the value of $k_i(T_k)$, the cumulative number of new ties, at the beginning of Ω increases rapidly as tie interactions with alters that existed prior to Ω appear for the first time within Ω. In our data where there is no left-censoring, this phenomena is picked up by the fact that the cumulative count of the number of new ties represents both a person's effective degree and new tie formation above and beyond their degree as noted in equation 10.

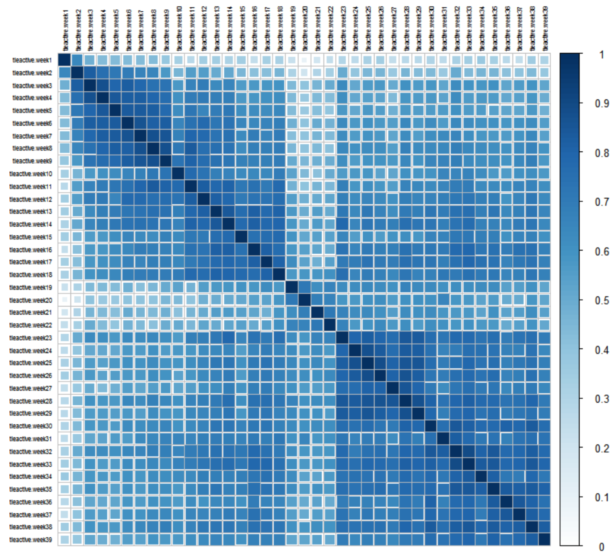

Figure 4: Inter-temporal Spearman Rank Correlations of Active Degree (κ_i^*)

11.5 ($\kappa_i^*(t_2 6, t_3 6)$ of the 28.5 ($\kappa_i(t_2 6, t_3 6)$ students that they could interact with during that period (41%).

4.4 Effective Degree Distributions

We have seen that the size of people's networks as measured by effective degree is very stable after an initial period of rapid tie formation. However, the standard deviations in Table 1 reveal that there is a good deal of variation across persons in the size of their networks. To explore this variation we decompose the total variation of effective degree into two component: variation within subjects and over time, and variation between subjects. Results reported in Table 3 show that virtually all of the variance in effective degree is between subjects and virtually none of it is over time (compete columns

	Mean	St. Deviation	Between Subject	Within Subject	Intraclass Correlation (rho)
Weeks 13-18	25.09	13.9	13.78	1.81	0.984
Weeks 26-36	28.49	16.05	15.91	2.06	0.984
Weeks 56-65	28.35	18.84	18.68	2.48	0.983

Table 3: Decompositon of Variance Within and Between Subjects at Three Points in Time

"Between Subject" and "Within Subject"). The intraclass coefficient (rho), a measure of how stable a quantity is, indicates that effective degree is so stable that its value in a given week within a semester can predict the value in any other week in that semester. Given this finding, the important variation that needs to be explored is not variation across time in a person's effective degree. For the most part effective degree is highly stable. What we need to focus on is the variance across person's in their effective degree.

This variation is depicted in the three figures in Figure 5 which show the effective degree distribution in the Fall 2011, Spring 2012, and Fall 2012 semesters. As noted earlier, the mean value for network size increases from Fall 2001 to Spring 2012, but remains that same in Fall 2012. However we see that there is a good deal of variation around these means. Furthermore, this variation appears to be increasing over time (see Table 1).[15]

Changes in the degree distribution are explored more by looking at the change score distributions in Figure 6. From Fall 2011 to Spring 2012 the mean change score is 3 with a standard deviation of about 5, indicating that the size of some people's networks declines while for others it increases. As

[15]There also appears to be changes in the shape of the distribution with some relatively large degree persons appearing in Fall 2012, and some skewness to the left.

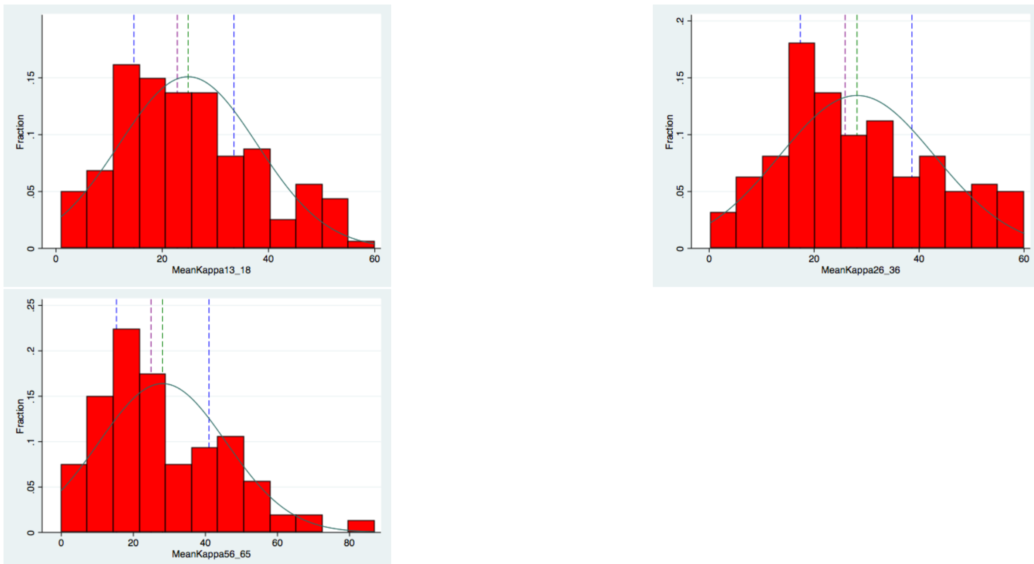

Figure 5: Effective Degree Distribution for Three Periods: Fall 2011 (Weeks 13-18), Spring 2012 (Weeks 26-36), and Fall 2012 (weeks 56-65), n = 161

noted earlier, from Spring 2012 to Fall 2012 there was very little change in mean level of effective degree, but there was still a good deal of variation around the mean change score (standard deviation of around 7). Clearly the degree distribution is spreading out as evidenced by both the increases in the standard deviation from Fall 2011 to Spring 2012, and the variance in the change score distributions which indicate that for some people the size of their network is increasing while for others it is decreasing. So even though the mean size of people's networks is fairly stable and a student's place in the degree distribution over time does not change very much, the "social" distance between people (as represented by differences in effective degree) is increasing.

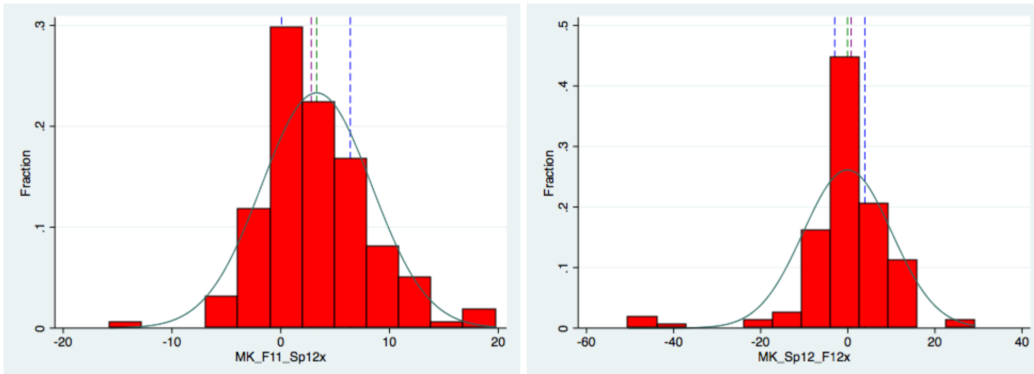

Figure 6: Distribution of Change in Effective Degree from Fall 2001 to Spring 2012, and from Spring 2012 to Fall 2012 NOTE on Ns

This increasing variation in the degree distribution over time is one sign that preferential attachment [REF] could be at work. According to the PA model, overtime those with higher degrees experience growth in their network size, while those with lower degrees experience decline either absolutely or relative to those with higher degree. According to this model, degree at time t should predict change in degree from time t to time $t+1$.

Figure 7 plots the change in degree from Fall 2011 to Spring 2012 by degree in Fall 2011, and the change in degree from Spring 2012 to Fall 2012, by degree in Spring 2012. There is clearly a small positive effect of degree at time t on change in degree from t to time $t+1$. The estimated regression line for both periods indicates that a one unit difference in effective degree yields a .1 change in degree, not a very large effect, but a significant one. Further inspection of these plots indicates that those with lower degree are more likely to experience a decline in degree (i.e. values of the change score less that 0), while those with higher degree are more likely to experience an increase in degree, though there are cases of high degree persons experiencing a decline in the size of their networks and low degree persons experience increases. The fact that degree at time t predicts to some extent degree at time $t + 1$, is part of the reason why there is stability over time in people's place in the degree distribution. However what is driving this system is the degree distribution in the Fall 2011 semester. PA amplifies the initial variation a the rich get richer and the poor poorer, but it does not alter people's place in that distribution. That place is determined early on in the process and thereafter is highly stable.

What is most striking about these degree distributions is, therefore, not how they are changing in shape over time, but that there is so much variation in effective degree from the very beginning, variation that is preserved over time. As noted earlier the variation between persons is substantially larger than the variation within persons over time. Students vary in their capacity to have networks of various sizes. Differences in this capacity are evident very early in the process. By the end of 12 weeks most students have reached their capacity as seen in the substantial differences across persons in their effective degree. Afterwards, effective degree is very stable as seen in the fact that there is little change in degree and in the fact that what change there is is predicted by the degree a person has after 12 weeks.

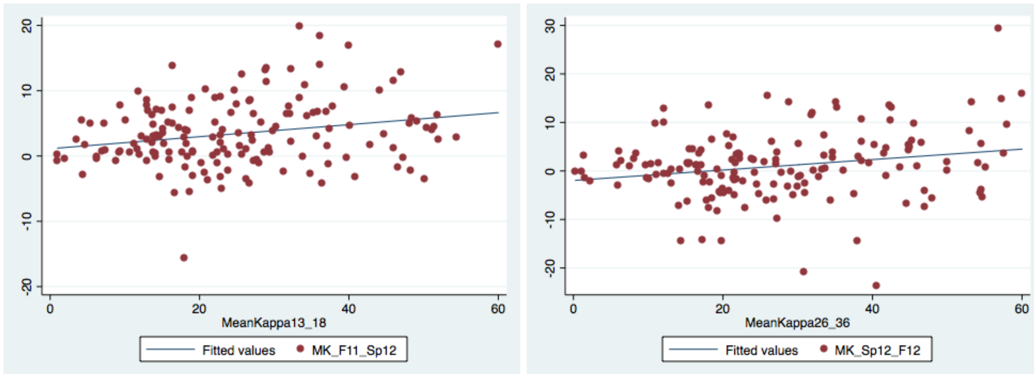

Figure 7: Scatter Plot of Change in Effective Degree Fall 2011- Spring 2012 by Degree in Fall 2011, and XXXX NOTE on Ns

5 Conclusion

The degree of persons and the degree distribution for a social network are well-known properties of social networks. By bringing a dynamic perspective to understanding degree we are able to distinguish and compare three conceptually and empirically distinct measures of degree: effective degree, cumulative degree and active degree. Effective degree captures the size of the set of alters that an ego *could* interact with at an given moment in time, cumulative degree the size of the set of alters an ego *has* interacted with during some time period, and active degree the size of the set of alters that an ego *is* interacting with in a time period. Effective degree best captures the notion of network range. As we have seen, it is relatively stable over long time-scales but varies a good deal across persons, an indicator that we should view effective degree more like an individual-level capacity. Individuals vary in their ability to create and maintain ego networks of various sizes. Some persons are more able to maintain larger networks, others have a difficult time doing so and learn to restrict the size of their networks accordingly. As a capacity, effective degree should be relatively stable over time because once a person reaches their capacity, forces hold in check both increases and decreases in network size for that person.[16] As such effective degree is closest conceptually to Dunbar's fixed inter-individual "ceiling" of sociability. This Dunbar number is not a constant within a population. People vary in their ability and desire to have larger networks. But as an individual level capacity effective degree is relatively fixed over time because people try to maintain a network of a given size with which they are comfortable.

Cumulative degree has often been thought of in the social network literature as capturing this capacity, but as a dynamic perspective shows, it conflates capacity with churn. The number of people an ego has interacted with during some period of time is a function of both the size of a person's network (i.e. the people they could interact with at any point in time) and the extent of churn in their network. Maintaining a network with a size that is compatible with a person's capacity requires replacing alters with other ties when relationships end and deactivating ties after forming new ties. The extent to which individuals in given periods of time deactivate ties (and then need to replace them) and activate ties (that require dissolving other ties)

[16] The assumption is that people become unhappy (or encounter problems) when the actual size of their network goes above as well as below their network capacity. When above, individuals will have difficulty maintaining all their relations. When below, individuals will feel that they are alone and seek out new relations.

determines the level of churn they will experience in their network. Individuals with similar capacities (effective degree) can vary in how much churn they experience. Those who have more churn will have higher cumulative degrees at the end of the time period, even though they have similar effective degrees. As such cumulative degree indexes a person's "history" of sociability which is a function of both their capacity to maintain a network of a given size and turnover within that network.

Active degree captures how much an ego interacts in a given time period with the alters with whom they could interact. Persons with similar effective degree may interact with their alters more or less frequently. Those who interact with them more frequently (i.e. have strong ties) are more active with their network than those who interact with the same number of people but on a less frequent basis.[17] As such, controlling for degree, more active persons will show up as persons who interact with more of their network within a given timeperiod. These discrete time activity levels can vary both across persons and over time. Some people may have more energy and time to strengthen their relationships with their alters. There can be times of the day, week, month, or calendar year in which those with ties to others interact with those others, and other times when activity levels within a population are very low.

Effective, cumulative and active degree measure different things and are

[17]NOTe: this is capturing the mean level of activity . Controls for degree. It also hides variation in how this activity level is spead out across alters

conceptually distinct. As this paper has argued, effective degree is the critical quantity. It captures the size of the pool of people a person can interact with, and as such it can be used to normalize other quantities. Churn should be measured not as the number of new ties, but that number relative to the size of a person's network. Similarly active degree is best conceived as the proportion of a person's effect degree pool that an ego interacts with in a given time period.

Furthermore, effective, cumulative and active degree have their unique "time signatures." Effective degree when it does change appears to change at the beginning and end of semesters, while active degree drops when students are off campus and increases during certain weeks within a semester. Churn seems to have less period effects and to be more episodic and less predictable.

Our "natural experiment" in which students with no prior ties to other students are placed on a campus and we observe the ties they form provides an ideal setting to observe the evolution of people's social networks and measure these qualities. As one would expect there is a great deal of tie formation in the first 12 weeks, but there is also some tie dissolution. What is most striking is that at the end of 12 weeks there is substantial variation in the size of students' networks *and this variation persists over time as does a student's position in the degree distribution*. This is all the more remarkable given that throughout the next 60 or so weeks, students experience a good deal of change in their networks. They form new ties and dissolve old ones. Periods of churn occur form time to time. Activity levels go up and down as students go on break and return to campus. Yet throughout all these changes their level of effective degree remains fairly constant. Effective degree appears to be something that persons bring with them when they enter new environments and form social ties. While its value is affected to a small degree by endogenous process such as PA, most of the variance in degree emerges very early on in the process. People bring to their networks a capacity to form and maintain ties, and this capacity varies across people. As they go about forming and dissolving ties, this capacity shapes their social network activity by restraining or promoting their desire to form ties, and by leading people to dissolve ties when they have more ties than they can handle.