Theory, Method and Data

Fred Stutzman

Table of Contents

Introduction

The study of transition and transition-related processes pose a number of important challenges to researchers. For example, the concept of transition is multi-dimensional (some transitions occur on a rang of levels, many occur simultaneously), and the general concept of a transition is not well-specified (disciplines employ varied understandings of the concept; within-disciplines, there is disagreement). In the following essay, I elaborate some of the challenges associated with studying transition-related processes, and explore research methods for analyzing and understanding the challenges. Finally, I explore a method, Structural Equation Modeling (SEM) that may be particularly useful for analyzing the multi-dimensional concept of transition.

Methods for Analyzing Transition

Among the many challenges of studying transition-related processes, one of the most important is specification. Therefore, I present the general model that guides this research. I argue that transition represents an interruption; the interruption created by transition is dual-process, with disturbance occurring at both affective and cognitive levels. The general process of accommodating the interruption of transition is known as adaptation; we are challenged to adapt to transition. As the interruption of transition is dual-process, so is the process of adaptation. The cognitive process of adapting to transition is known as integration. In integration, the individual adopts new roles and identities, addresses informational challenges, and develops and maintains supportive social networks. Transitions have been demonstrated to be correlated with stress symptoms, and therefore the affective challenge of transition is managing stress. It is argued that stress is both buffered and impacted through the delivery of social support. Social support is a variable construct describing the support provided by one's social network. General components of social support include emotional, tangible, and informational support. Furthermore, this support can occur at the perceived level (in which the individual feels supported) and at the received level (in which the individual actually calls on the network for support).

Given that adaptation to transition is dual-process in nature, a range of theoretical approaches have emerged to describe adaptation. The cognitive processes of adaptation are generally described by process theories, such as those by Ebaugh, Ashforth, and Nicholson. The affective processes of stress management are generally described by variance theories, such as those by Sarason and Sarason, Scholssberg, and Cohen. Process and variance theories have fundamental differences; whereas variance theories are amenable to hypothesis testing, process theories are useful for specifying lenses for the analysis and integration of findings. In the study of transition, process theories usually emerge from grounded inquiry (e.g. Ebaugh), and are therefore explored with qualitative methods (primarily in-depth and semi-structured interview). Variance theories, on the other hand, are generally tested with observational (e.g. survey) and laboratory data. The majority of studies on transition employing variance theories use survey data and statistical techniques to test hypotheses.

The decision to employ interview and survey data in the proposed research, therefore, fits existing (process and variance) approaches to analysis of transition. These two data sets are complementary, allowing for within-population analysis of the process and outcomes of social network site use during a transition. In addition to this data, I specify the use of a third dataset of in-transition Facebook profiles to facilitate two additional approaches to the study of transition.

Particularly, I propose to study the dynamics of support networks (e.g. networks articulated in social network sites) through network and longitudinal models.

To argue the value proposition of the network models, I first specify the value propositions of the two other data sets.

- Interviews allow the examination of the process of transition, facilitating an understanding of how the social network site addresses the process of transition.
- Survey analysis allows an understanding of outcomes associated with use of the social network site during transition.

Between these two data points is a gap, which I've identified as the dynamics of support. Therefore, it is not enough to simply know how the sites are used, and correlates of use (i.e. outcomes), we should also investigate the dynamics and flows of support through the network. The analysis of dynamics allows the bridging between process and variance theories of transition; the approach I apply is similar to the work done by Burt in the book *Structural Holes*. In the book, Burt argues that the use of network and statistical method can identify the "central" individuals in the network: the brokers through which support and information flow. Therefore, by integrating these three approaches to analysis, I believe I develop a methodology that provides greater nuance by bridging the process and variance approach, and provides a roadmap for future analysis.

To the question of opportunism, I can't deny that there are opportunistic aspects of the analysis. The choices of analysis fit those that I know and have practiced. The data collection strategies are also those I'm familiar with. I also wish that I could have up-to-date social network site data, but unfortunately I am only allowed to use data up until mid-2006. However, research is always a tradeoff between quality and validity, and I believe that combining the three forms of analysis provides the greatest "bang-for-the-buck" possible through the combination of large samples and qualitative analysis.

Further, the data sets I have selected are, to the best of my knowledge, those most appropriate for the questions I've specified. Process models of transition are studied and developed with qualitative interviews, whereas variance theories are most often tested with survey data. The analysis of network dynamics is particularly amenable to longitudinal network data; in this sense, I have an extraordinarily rich data set. Finally, ability to bridge the gap between process and variance theories through network analysis represents an important contribution, both to the richness of this project and for other researchers interested in this topic.

When arguing the validity of this integrative approach, one question I expect is "what does a Facebook network actually tell us about support?" In my process and variance modeling, these connections are fairly straightforward (e.g. using validated scales and constructs), but less so in the dynamic model. What is the value of a Facebook friend?

The analysis of networks takes two primary forms: networks and graphs. The primary difference between networks and graphs is that networks have "value" assigned to the ties between nodes. In this sense, Facebook represents an unstructured graph, where all ties are equal in the socio-technical system. To understand ties we can apply the fairly standard construction of tie strength. Strong ties are associated with bonding support and capital (i.e. strong in-group support, high reciprocity and trust), whereas weak ties are associated with bridging support and capital (i.e. ad hoc in nature, low reciprocity). Although counterintuitive, weak ties have been demonstrated to be valuable in economic situations (e.g. getting a job), and it is has been demonstrated that increased weak ties produces greater bridging social support and capital.

Within the studies of social support in online systems, researchers often encounter unstructured graphs (e.g. Bambina). Some researchers have attempted to derive tie strength behaviorally (e.g. Gilbert), but many studies rely on unstructured graphs, assuming that more ties equals more support. Although commonly employed, I think this is a weak assumption. Let us briefly consider the functional form of social support:

$$f(ss) = a + b(received) + e \quad \textbf{or} \quad f(ss)\ a + b(perception) + e$$

Theories of social support generally argue that social support acts as either a buffer or direct effect in the reduction of stress. Therefore, we can define the relationship between stress (s) and social support as:

$$f(s) = a + b(ss\text{-}received) + b(ss\text{-}perceived) + e$$

In this functional form, we see that stress is a function of both perceived and received social support. This model assumes interaction; to perceive and receive support one must interact with others. So what is different about a social network site? If support in a social network site was purely a function of active interaction, I'm not sure how interesting or useful they'd be in context. However, social network sites contain activity indicators and stimulus generating features such as the News Feed. In a social network site, ones perception of the network is a function of the network articulated, individual interaction, and stimulated interaction. Therefore, by having larger networks, one may be subject to greater and more diverse stimuli, thus enhancing the supportive perception and capacity of the network. The functional form of support in a social network site is then:

$$f(sns\text{-}support) = a + b(network\ size) + b(frequency\ of\ use) + b(diversity\ of\ news\ feed) + b(interpersonal\ interactions) + e$$

Accordingly, the general relation between support and stress can be specified as:

$$f(s) = a + b(ss\text{-}received) + b(ss\text{-}perceived) + b(sns\text{-}support) + e$$

Where sns-support is a function of the size, diversity, and resources available within the social network site.

Challenges of Integration and Side Effects

In the previous section, I highlighted some of the theoretical and methodological challenges associated with analyzing transition. In this section, I explore some of these challenges in greater depth, at the theoretical and functional level, and then specify some strategies for managing side effects.

In the study of social network sites, and behaviors therein, we must consider the contexts of the data collected. That is to say that the data we can derive from the study of networks has many forms and limitations. boyd and Hogan outlined some of these challenges in their essay on the multiform nature of networks. The authors argue that networks in social network sites are represented on three levels: behavioral, ego, and articulated.

Behavioral networks refer to the networks created by interaction in the social network site. For example, this network could contain our most frequently emailed contacts, or the profiles we view the most, and so on. The data points we leave behind allow systems to create representations of networks that can be used for relevance judgments, privacy features, and so on. Ego networks are in line with the sociological sense of "social networks." Ego networks represent who we know or are connected with in any way. Therefore, social network sites and ego networks are almost always exclusive; some people in our social network won't be members of the social network site (this may change, but for now I'd argue it is valid). Finally, articulated networks refer to the Friend networks we create in social network sites. These networks are curated presentation of a social network, they are ties that we wish to articulate.

It is important to remember limits to the validity of judgments we can make from these different views on networks. Recently, Google introduced the Buzz social network platform. Buzz investigated Gmail behavioral networks, and pre-populated an articulated network (i.e. a Friend list) based on the behavioral network. This caused a major privacy uproar, as people think of behavioral and articulated networks differently. The people we email the most (e.g. people at the compliance department) do not necessarily comprise the networks we'd wish to articulate. I think this raises an interesting question: what is the true network? Adopting a symbolic interactionist perspective, I would argue that each are valid on some level, none are more real that the other. The management of these networks represent an everyday process, and when technologies expose one network (e.g. behavioral) in a place where we'd rather have a curated network displayed, there exists the potential for validity errors and privacy risks.

At a functional level, I've tried to answer this challenge through mixed-method analysis. In my proposal, I propose analysis that allows both within and between comparisons in my data sets. My qualitative and survey data can be compared within, and the network and longitudinal models can be compared within. I can then compare between these sets (though only generally as the data sets are from different times). I believe this approach allows both a process and variance theorizing of the use of social network sites during transition, and the analysis of dynamics represents middle ground between the process and variance theories. For example, network site is often shown to be associated with positive outcomes in variance models. By providing insight into the dynamics of network growth, I can explore factors associated with growth that would not exist in survey data. Through the use of comparisons between my important data points, I hope to be able to put them in context and get a better sense of their limitations.

With regards to managing side effects, I first state that methodological integration is challenging. We often see papers that employ both survey and interview data, in which the interviews are targeted towards particular interesting features of the survey. Such an approach stands to potentially bias the surveys, as the questions have been fit to a prior that may have only limited validity. In my analysis plan, I approach the surveys and interviews independently, with both question sets ground in theory; the comparisons I make between will come after (as opposed to being led by) the analysis.

Notably, the process of "fitting" questions to interesting findings can be informative, so I do not completely dismiss it. Finding negative cases, for example, is a valid process for identifying where variance models fail. However, my goal in the process is to develop data sets that are mutually informative, and enable between comparisons so I can identify weaknesses with the approaches I've considered.

With regards to the network data I've collected, I would argue that due to the observational nature of the data, it may be less challenging to compare between the analyses. Employing a structural holes approach, I use a range of tactics to analyze the limitations of the particular configurations of the networks. One particular limitation may be that network analysis is exploratory in nature. While hypotheses can be tested through such methods as Monte Carlo simulations, I will be doing my hypothesis testing through more standard analytic techniques.

A final challenge of network data deals with its validity. Although I've previously explored the value of network ties, and the limitations of behavioral networks, I must remember that network data is but one component. Although it has many desirable qualities, there are many limitations of behavioral networks (cf. Google Buzz), and I must not bias my opinions too strongly toward this data-rich set.

SEM and Multiple Data Points

Structural Equation Modeling is a form of linear modeling that allows for the simultaneous testing of the direction and effect of observed and latent variables, as well as the correlation of error terms between these models. Using regression and confirmatory factor analysis, SEM allows for hypothesis testing in complex, simultaneous models. Furthermore, SEM allows for the estimation of model fit through a range of means, including the chi square, RMSEA, and others.

In the proposed research, I plan to use SEM to evaluate the relationship between social network site use, social support, stress and adaptation, as well as a range of empirically and theoretically relevant covariates. SEM represents an attractive analysis technique for a number of reasons. First, SEM is informative for complex models, such as the one I propose. Through the use of simultaneous hypothesis testing, I will be able to identify particularly important components related to the theoretical models specified. Second, SEM allows the analysis of unobserved error. For example, if an unspecified latent construct is creating patterns of error covariance in my models, SEM will identify this. This knowledge can be used for fine-tuning or re-specification in future research. Finally, SEM provides estimates of model fit that will allow me to compare between theoretical processes specified, providing justification for best-performing models.

To this point, I have primarily discussed how SEM is useful within a data set (particularly, my survey data set). How could SEM be useful in a mixed-methods analysis? I propose three methods. First, assuming proper specification and identification, the SEM model will provide guidance on where to "go" with a mixed-methods analysis, particularly identifying what sub-components of the process have strong effects, or are counter-intuitive. The detailed analysis allowed in SEM will enable a targeted, in-depth investigation of my other data sources in this case.

Second, the SEM's ability to model correlated errors is particularly important, as it may identify third-party processes that are influencing the use of social network sites for social support or adaptation. An example might be an external or structural force that may be reducing adaptation. In this case, I could then investigate the qualitative or network data to develop new hypotheses regarding the correlated error. Third, SEM is particularly good fit for the analysis of latent constructs, in situations where error is significant (i.e. surveys, self-reports). SEM is useful as a rich explication of process, facilitating following longitudinal analysis. Therefore, SEM is not only useful within my current data scheme, but it sets up following analyses well.

Conclusion

In the previous essay, I elaborate some of the challenges associated with studying transition-related processes, and explore research methods for analyzing and understanding the challenges. I then explore a range of processes through which social support is provided in social network sites. I then leveraged a range of perspectives on networks, highlighting the limitations and benefits of my mixed-methods analytic approach. Finally, I explore a method, Structural Equation Modeling (SEM) that may be particularly useful for analyzing the multi-dimensional concept of transition.

Appendix: Question

Day 5. Theory, Method, and Data

You bring together a range of theoretical models to inform your research questions and say that you aim to develop a methodology for studying information behavior during life transitions. Your approach outlined in the introduction is a mixed method attack that uses three snapshots through 3 different lenses (logged behavior, large-scale survey, and in-depth interviews) over two distinct time periods.

1. How does this particular combination of data and methods define a new methodological approach that is tuned to the relationships among social support, stress, and social networks during university transition? [why is this not simply an opportunistic combination of data that you already have with some new things?]

2. What are the main methodological challenges of integration of results in mixed methods approaches?

3. For each of these challenges, what can be done to ameliorate side effects?

4. How might structural equation modeling (SEM) help in integrating multiple data streams?

When possible, use examples from your literature such as how do the differences in perspective among different models affect these challenges (e.g., Schlossberg's context dependent model versus Nicholson's more general process model; information seeking models and motivations versus information sharing models and motivations).