Wearable Bluetooth sensors for capturing relational variables and temporal variability in relationships: A construct validation study

James G. Matusik[1]

Ralph Heidl[2]

John R. Hollenbeck[1]

Andrew Yu[1]

Hun Whee Lee[1]

Michael Howe[3]

[1]Michigan State University
[2]University of Oregon
[3]Iowa State University

Abstract

The advent of wearable sensor technologies has the potential to transform organizational research by offering the unprecedented opportunity to collect continuous, objective, highly granular data over extended time periods. Recent evidence has demonstrated the potential utility of Bluetooth-enabled sensors, specifically, in identifying emergent networks via co-location signals in highly controlled contexts with known distances and groups. Although there is proof of concept that wearable Bluetooth sensors may be able to contribute to organizational research in highly controlled contexts, to date there has been no explicit psychometric construct validation effort dedicated to these sensors in field settings. Thus, the two studies described here represent the first attempt to formally evaluate longitudinal Bluetooth data streams generated in field settings, testing their ability to (a) show convergent validity with respect to traditional self-reports of relational data, (b) display discriminant validity with respect to qualitative differences in the nature of alternative relationships (i.e., advice versus friendship), (c) document predictive validity with respect to performance, (d) decompose variance in network-related measures into meaningful within- and between-unit variability over time, and (e) complement retrospective self-reports of time spent with different groups where there is a "ground truth" criterion. Our results provide insights into the validity of Bluetooth signals with respect to capturing variables traditionally studied in organizational science and highlight how the continuous data collection capabilities made possible by wearable sensors can advance research far beyond that of the static perspectives imposed by traditional data collection strategies.

Keywords: wearable sensors, Bluetooth, convergent validity, predictive validity, network dynamics

Organizations are complex social systems, composed of networks of individuals that interact, communicate, and coordinate activity within and across levels and time (Kozlowski & Klein, 2000). Accordingly, a great deal of organizational research has focused on relational variables at various levels of analysis, including the dyadic (i.e., two individuals), team, and network (i.e., whole system) levels. For example, hundreds of studies have examined the important, dyadic relationships that form between leaders and followers (Dulebohn, Bommer, Liden, Brouer, & Ferris, 2012). Likewise, teams research has been recognized as a dominant and growing topic area in organizational science (Mathieu, Hollenbeck, van Knippenberg, & Ilgen, 2017; Morrison, 2010). In the domain of network research, network characteristics have been identified as major determinants of a variety of outcomes (Crawford & LePine, 2013), including knowledge creation, transfer, and adoption (Phelps, Heidl, & Wadhwa, 2012), career success (Seibert, Kraimer, & Liden, 2001), and socialization (Morrison, 2002). Indeed, the recognition that organizations are inherently relational arenas has propelled organizational science forward.

Some limitations associated with methods widely used in organizational science, however, have thus far hindered researchers' ability to capture the richness of many relational constructs, especially as they change or evolve over time. Generally speaking, research on relational variables has relied on cross-sectional or before/after approaches captured with self-reported data or direct observations from others, both of which are affected by the limits of human memory and subject to a host of biases (Lazer et al., 2009). Furthermore, traditional approaches, such as surveys and diaries, are taxing on participants and often result in low response rates, especially in longitudinal contexts where repeated measurements are required. Even if one moves beyond surveys and diaries, other existing methods of studying human

interaction, such as video and audio recordings, are time consuming to code and impose geographical constraints on data collection.

Fortunately, recent technological advancements have the potential to overcome some of these limitations, supplement traditional data collection practices, and, consequently, open entirely new lines of research. Specifically, wearable sensor technologies (WSTs) could revolutionize the way organizational scientists conduct research on a range of relational variables (Pentland, 2012). The use of WSTs is already gaining traction in many applied contexts where they have been used to track interactions among individuals in a variety of environments, including schools (Fournet & Barrat, 2014; Mastrandrea, Fournet, & Barrat, 2015), social events (Panisson et al., 2012), and even work organizations (Alshamsi, Pianesi, Lepri, Pentland, & Rahwan, 2015). Despite the proliferation of WSTs, however, there has been little in the way of traditional construct validation studies dedicated to measures derived from these sensors. Indeed, construct validation evidence for assessing relational variables with wearable sensor data is scarce for controlled environments (for an exception see Chaffin et al., 2017) and non-existent for extended duration field studies.

This is problematic because WSTs must be shown to be construct-valid data collection instruments (Schwab, 1980) before researchers use them to conduct substantive organizational research and before employers use them to track behavioral patterns. Hence, the primary purpose of the studies presented here is to validate one particular WST, namely Bluetooth (BT), in field settings. We chose to focus exclusively on BT given the results of Chaffin et al. (2017), in which it was determined that, among several WSTs examined, the most promising and robust WST was the BT sensor. As such, this current investigation is in no way the first "ground truth" assessment

of BT, which was provided by Chaffin et al. (2017) in their series of laboratory studies. Instead, it is the first realistic, multi-day validation effort of WSTs in field contexts.

In our first of two studies, we use wearable BT sensors (manufactured by Sociometric Solutions, recently rebranded as Humanyze) to capture interaction dynamics in a large-scale, multimillion dollar scientific installation. In the second of two studies, we use wearable BT sensors (manufactured by Limefy) to track day-to-day interactions among team supervisors in a heavily monitored environment, providing additional ground truth evidence but in a less controlled setting than those in Chaffin et al (2017). In conducting this second study, we not only evaluate the validity of this alternative BT sensor but also assess its convergence with the BT sensors embedded in the Humanyze device. The results of this test of convergence support Chaffin et al.'s (2017) recommendation and our own belief that researchers should focus on the WST itself (i.e., BT), such that the technology should not be confounded with its instantiation (i.e., the manufacturer).

Thus, this validation effort builds upon the laboratory studies of Chaffin et al. (2017) with two new field studies, and it represents the next logical step towards establishing WSTs as valid data collection instruments. Importantly, the three most basic differences between this study and Chaffin et al. (2017) are that (a) Chaffin et al. (2017) was based entirely on laboratory research that was (b) specifically aimed at assessing the ground truth accuracy of several different WSTs (BT, microphones, and infrared detectors), and (c) did not capture relational or network dynamics as a function of changes in activity states in a field setting. While laboratory research is an important endeavor and critical first step before WSTs are deployed in field contexts, evidence of their validity in field contexts could add substantially to their perceived utility by practitioners and organizational researchers alike. This is evidence we hope to provide.

To accomplish this objective, we evaluate the construct validity of BT sensors using traditional psychometric procedures, assessing their ability to capture interactions as well as fundamental relationships often studied in organizational research (i.e., advice- and friendship-oriented relationships). Specifically, we test the ability of BT-generated data to demonstrate (a) convergent validity with respect to traditional self-reports of relational variables, (b) discriminant validity with respect to qualitative differences in relationships (i.e., advice-oriented versus friendship), (c) predictive validity with respect to individual performance in a context where boundary spanning (i.e., brokerage) is critical, (d) systematic variance decomposition of within- and between-unit variation in common network constructs at the "daily" level, and (e) incremental validity, beyond retrospective self-reports, for predicting time spent with different groups in a context where ground truth is known.

The results of this validation effort offer several contributions to organizational research. First, we provide evidence that BT sensor-generated proximity data may be able to detect and track affective and task-oriented relationships among individuals in field settings. Indeed, our findings suggest that BT sensors can assist researchers interested in studying friendship and advice-oriented relationships, which have been identified as central elements of informal organizational networks (Gibbons, 2004; Morrison, 2002), as well as established network characteristics such as brokerage opportunities (Burt, 2004). Second, we demonstrate that BT sensor-generated data is capable of capturing interpersonal interactions, more generally, as we provide evidence that this data may be used to determine the proportion of time one spends with others in various, known locations, above and beyond self-reports. By providing evidence of wearable BT sensors' ability to estimate known time spent with others in field settings, as well as capture self-reported relationships, we contribute to the realization of their potential as a valuable

addition to existing methods of data collection that is (a) less time intensive and vulnerable to human subjectivity than traditional methods of data collection, (b) deployable on a large scale and over extended time periods, and (c) capable of facilitating the fine-grained, longitudinal examination of interaction patterns among individuals at an unprecedented level (e.g., hourly, daily, weekly).

Third, and importantly, we find that the exploitation of this technology's unique strengths may allow researchers to address entirely new research questions. Above and beyond their ability to provide more objective and efficient methods for data collection, the temporal granularity associated with wearable BT sensor-generated data opens novel lines of research that could hardly be contemplated, let alone attempted, with existing methods. Specifically, the opportunity to more deeply explore issues related to the dynamic nature of individual and organizational network characteristics, with a focus on "within-unit" rather than "between-unit" variability in network variables (i.e., how one's position in the network varies on a day-to-day basis rather than how one's typical position in the network differs from the typical position of other individuals, more generally), has been difficult to explore. This is due almost entirely to the limitations associated with traditional, labor intensive survey approaches that tend to rely on a reduced number of coarse-grained, aggregate snapshots that make the implicit assumption that relationships among individuals are relatively stable and largely focus attention on between-unit variance. If recent trends in organizational research have anything to teach us, it is that unpacking within- and between-unit variance often leads to profound discoveries.

Indeed, the importance of isolating and separating within- versus between-unit variability has been documented repeatedly in various subfields of organizational research (e.g., Dalal, Lam, Weiss, Welch, & Hulin, 2009; Matta, Scott, Colquitt, Koopman, & Passantino, 2017; To,

Fisher, Ashkanasy, & Rowe, 2012), at times casting doubt on the meaningfulness of the between-unit score. At the very least, the recognition that a large percentage of variance on a given variable was residing at a heretofore unrecognized level opened new avenues of inquiry, such as why a leader may adhere to justice norms on one day, but not another (Scott, Garza, Conlon, & Kim, 2014). In the present investigation, we show how wearable BT sensors, which scan their environments multiple times per minute, offer the unprecedented opportunity to unpack within- and between-unit variance in network characteristics at multiple levels of analysis. Our analyses show that a significant proportion of variance in some individual network characteristics resides within individuals, such that any single individual's position within the broader network may vary greatly from day-to-day. Furthermore, we provide evidence that this within-individual variance is meaningful as it may reflect actual changes in activity states of the entire network under examination (e.g., the network's transition from planning a project to actual execution).

Finally, we contribute to computational social science research on organizational behavior (Olguin et al., 2009; Pentland, 2012) and social network dynamics (Cattuto et al., 2010; Sekara, Stopczynski, & Lehmann, 2016), more generally, by providing a comprehensive validity assessment of a frequently used data component (i.e., BT). Computational social science is fundamentally data-driven and leverages the increasing availability of large scale, longitudinal data that contains relational information, such as phone call records, emails, GPS, and radio frequency identification data, to study individual and collective behavior (Lazer et al., 2009). In doing so, computational social science researchers often triangulate BT data with data from other sources and sensors to increase the validity of derivative relational variables (Stopczynski et al.,

2014). Thus, our research complements and informs computational social science research on social networks by providing evidence concerning the validity of BT-derived relational variables.

Wearable Sensor Technologies in Organizational Research

Wearable sensor technologies (WSTs) hold a great deal of promise for organizational research because they allow researchers to collect copious amounts of objective, granular data over extended time periods. However, a major weakness of WSTs often overlooked by researchers is that each sensor must be attached to an individual, and alternative attachment modes strongly influence the accuracy of measurements. That is, several WSTs may be combined into a multi-sensor, wearable measurement system that has a single point of attachment (e.g., worn loosely around the chest), and that single point of attachment may not be optimal for the different component sensors. For example, the audio signal produced by a microphone is compromised if it is captured at a person's chest instead of near his or her mouth. The light detection signal of an infrared sensor, interpreted as face-to-face alignment, is similarly compromised if the sensor is placed on the chest and not the face. Indeed, in assessing the validity of multi-sensor wearable measurement systems, Chaffin et al. (2017) determined that variance exists in the usefulness of various sensors, most of which could be traced to attachment problems and differences in sensitivity between microphones, infrared sensors, and accelerometers.

One WST examined by Chaffin et al. (2017), however, offered greater proof of concept relative to others. Specifically, Bluetooth (BT) sensors were determined to be very accurate in detecting proximity events over extended periods of time where ground truth was known. Furthermore, BT sensors were far more robust proximity sensors than infrared with respect to errors of omission (i.e., false negatives). These findings, coupled with the ubiquity of BT in both

the emerging WST literature (e.g., Eagle, Pentland, & Lazer, 2009; Sekara & Lehmann, 2014; Stopczynski et al., 2014) and mobile consumer devices, led us to focus explicitly on wearable BT sensors in the present investigation, with the goal of assessing their validity and utility for organizational research.

Wearable Bluetooth Sensors

BT sensors are pervasive commodities that are standard components embedded in many consumer communication devices. Originally conceived as a data transfer technology (Ferro & Potorti, 2005; Haartsen, Naghshineh, Inouye, Joeressen, & Allen, 1998), BT sensors can detect the presence of other BT-enabled devices within a specified range, with the maximum range of detection ranging from 1 to 100 meters (Wang & Iqbal, 2006) depending on sensor specification. The minimal inquiry length (i.e., temporal resolution, or amount of time between sweeps for other sensors) of BT sensors varies between sensors, with some sweeping once every couple of seconds and others sweeping approximately once per minute (Kammar, McNutt, Senese, & Bray, 2002). Given their range and inquiry length, BT sensors can generate a considerable number of detections in a relatively short period of time.

A detection generated by a BT sensor includes information about (a) when the detection took place and (b) the strength of the received radio signal (i.e., how far apart the two sensors were when a detection was registered). Thus, BT data records consist of a time/date stamp, sender identifier, receiver identifier, and a *Radio Signal Strength Indicator* (RSSI). RSSI is a measure of the strength of the signal received from a proximate BT sensor. Importantly, the range of RSSI depends on the specifications of the BT sensor.

The ability of BT sensors to detect other BT-enabled devices has been leveraged by prior researchers to record and infer co-location situations among individuals (e.g., Eagle et al., 2009;

Sekara & Lehmann, 2014). In assessing the potential of BT technology for research, it is important to recognize that the original purpose of RSSI, which researchers often rely on to determine whether a registered BT detection represented a meaningful interaction opportunity, was to describe the quality of a wireless communication link for data transfer. The use of RSSI to measure spatial separation in human interaction is a relatively novel application that falls outside the original scope of the technology's intended use. Thus, a clear understanding of the relationship between BT RSSI and spatial separation is a necessary precondition for meaningful behavioral applications. In this context, three characteristics of BT RSSI are of importance.

First, RSSI is recorded as a negative value and typically ranges from -50 to -100 (though this range may vary by sensor). Less negative RSSI values, or values that are smaller in absolute value, represent shorter distances between BT sensors while more negative RSSI values, or values that are larger in absolute value, represent greater distances between BT sensors. Thus, a detection with an RSSI value of -70 presumably represents a shorter distance between BT sensors (when the detection was registered) than a detection with an RSSI value of -75, -80, -95, and so on.

Second, RSSI data recorded by BT sensors positioned at a fixed distance follow the pattern of a Gaussian normal distribution (Ramadurai & Sichitiu, 2003; Sichitiu & Ramadurai, 2004). This is important because it suggests that while BT detection strength (RSSI) may produce a noisy and relatively unreliable measurement of spatial separation for any individual observation, its random error distribution should allow the "true" RSSI value to emerge over several observations. This has implications for behavioral research designs concerning the choice of a suitable observation and deployment period for BT sensors, such that larger windows of observation should facilitate the cancellation of random error.

Third, because BT signal attenuation does not follow a linear path and depends on sensor specification, the traditional engineering practice of establishing a dichotomous, binary (i.e., yes/no) signal strength *threshold* to identify co-location events between two individuals is problematic, psychometrically. For example, previous sensor-specific calibration efforts identified an RSSI of -80 as an appropriate threshold (approximately 3 meters; Kalimeri, 2013) for face-to-face interactions (Finnerty, Kalimeri, & Pianesi, 2014; Olguin et al., 2009), such that any detections with RSSIs less than -80 (e.g., -81) would be excluded from analyses as they would represent distances supposedly too great to suggest interaction (thus, eliminating conference room interactions, as well as individual broadcasting to people in a small classroom). Indeed, recent evidence suggests that the choice of a particular threshold setting has a substantial impact on the derivative co-location structure (Chaffin et al., 2017). Thus, meaningful binary threshold-based co-location diagnostics are contingent on the appropriate definition of an interaction threshold distance (highly context-specific) and require the calibration of every BT sensor RSSI with respect to this distance.

Moreover, by using binary BT RSSI thresholds, one is transforming a continuous variable into a dichotomous variable and discarding potentially meaningful signal variance in the assessment of relational behavior (e.g., friendship versus task-related interactions). Although some researchers have found evidence that thresholding techniques might allow one to infer qualitative differences in interaction patterns (Sekara & Lehmann, 2014), others have acknowledged that using a relatively strict cut-off, such as one that corresponds to just a few meters (e.g., 3 meters), could lead to many false negative detections (Onnela, Waber, Pentland, Schnorf, & Lazer, 2014). The recognition of these technical and analytical limitations clarifies why initial attempts to leverage BT technology for social and behavioral research need to use a

wide range of BT-generated proximity data (in terms of RSSI) and collect this data over extended time periods.

Fortunately, a key advantage of WSTs is that they do allow for the mass collection of high-resolution, temporal data over extended observation periods, which may help to overcome some of the aforementioned limitations while also providing insight on the social context in organizations (i.e., relationships). Leveraging the full information contained in the RSSI signal avoids issues with determining an appropriate dichotomous threshold, reduces the restriction in range which may otherwise limit the ability of BT data to converge (i.e., correlate) with self-report relational measures as traditionally assessed in organizational research, and opens the possibility of discriminating between different interaction contexts. Furthermore, collecting this data over extended periods allows random error to cancel out and provides a relatively stable and accurate estimation of true proximity.

Equipped with an understanding of BT's usage in the burgeoning WST literature, as well as the information embedded in the data that BT sensors generate, we now explicitly turn our attention to the validation of this technology in field contexts.

Validation of Bluetooth Sensors

Traditional psychometric construct validation of BT-generated data is an important first step before any substantive research (or employer-based tracking) involving BT sensors should take place. Ensuring construct validity, or correspondence between relational variables (i.e., constructs) and BT-generated data (i.e., measures) (Schwab, 1980), is critical for at least three reasons. First, without validation we run the risk of mistakenly interpreting spurious findings due to the high-powered empirics associated with the large number of observations BT sensors generate. As noted, BT sensors are capable of generating sizable amounts of data given their

frequent sweeps for other BT-enabled devices, and therefore they provide researchers with a high level of statistical power. To ensure that meaningful patterns of interaction and activity are being captured by BT sensors, rather than mere co-location, it is critical that evidence of their ability to detect known relationships is provided in advance.

Second, failure to validate BT sensors before engaging in substantive research or practical use makes it impossible for organizational researchers to identify the reasons that they are unable to replicate their own or others' results. That is, without validating BT sensors and other WSTs before utilizing them, it is not possible to know whether a failed replication using BT-generated data is a product of the researcher's methodological approach, their theory, or the technology itself.

Finally, and as recognized by Chaffin et al. (2017), many WSTs offered by manufacturers provide constructs that are defined with engineering standards in mind and calculated using proprietary algorithms. That is, to date, evaluation standards for variables derived from WSTs have been approached from an engineering angle rather than a psychometric angle, and in many cases the goal that an engineer has when creating a metric differs from that of a psychometrician. For example, in an effort to measure social dominance an engineer's priority is to ensure that the minutes of verbal activity recorded by an audio device is identical to the actual number of minutes spoken. However, when constructing a measure of social dominance, a psychometrician might only be interested in multiple fallible indicators that capture the true rank-ordering or equal-interval discrimination of people on that broad construct. From a construct point of view, a valid rank-ordering or equal-interval discrimination of people from multiple indicators is more relevant in assessing social dominance than the simple count of minutes of speaking time (or some permutation thereof), which is likely to be contaminated and/or deficient for the purpose of

describing social dominance. In addition, a traditional practice within engineering fields is to use "thresholding" to convert continuous data into dichotomous counts, where in traditional psychometric contexts every effort is made to keep variables in their continuous form (e.g., Stöber, Dette, & Musch, 2002).

Although engineers and psychometricians strive to achieve similar goals via different methods, both groups often use the same vocabulary when referring to relational constructs. When one couples these differences with the fact that most WST providers rely on complex and often proprietary signal processing algorithms, this makes it impossible for researchers to evaluate these newly (re)defined constructs. Therefore, validating and utilizing raw detection signals (i.e., BT-generated data) may prove to be a more promising and successful approach than naïvely relying on proprietary, manufacturer-provided constructs, which is often the alternative when one uses WSTs. Accordingly, our attention is explicitly focused on the validation of the transparent, raw detection data generated by wearable BT sensors.

Convergent Validity of Bluetooth Data with Self-Reported Relational Variables

As noted, the ability of BT technology to detect other BT-enabled devices has been leveraged by prior researchers to record and infer co-location situations among individuals. Given potential concerns about the validity of this technology, our first research question pertains to the *convergence* of BT proximity data, generated in field contexts, with self-reported relational data. Specifically, we examine the extent to which BT data converges with self-reports of friendship and advice-oriented relationships as these informal ties have been identified as key elements of informal networks (Gibbons, 2004; Morrison, 2002). Although self-reports are not always completely accurate, BT sensors should generate data that converges with commonly

used measures in organizational research for them to represent valuable data collection tools in this domain.

This test of convergence represents the fundamental first step in our investigation as a lack of evidence of convergent validity would preclude the possibility of interpreting the results of subsequent analyses. Without first establishing that BT-generated data does indeed correlate with self-reported relational data, the results of all other analyses would be highly questionable. Thus, to take the initial step towards integrating BT sensors into organizational research, we use a wide range of proximity data collected over several weeks in a field setting to address our first research question:

Research Question 1: Does cumulative Bluetooth proximity data converge with self-reported relational data in a field setting?

Discriminant Validity of Bluetooth Data for Assessing Alternative Relational Variables

Our second research objective pertains to the use of BT proximity data to *discriminate* between various types of relationships. Whereas WSTs can help overcome some of the shortcomings of self-reports, the value of the data generated by proximity sensors may be limited by the absence of contextual information related to subtle distinctions between various types of relationships. For example, traditional network surveys make distinctions regarding the nature of interactions by offering different items for advice- versus friendship-oriented relationships. In contrast, current analytical approaches related to BT data consider all proximity detections above a single dichotomous threshold (e.g., RSSI of -80) essentially the same. If the nature of human interaction were to be reflected in spatial separation (i.e., personal space), alternative analytical approaches could exploit the richness of the granular proximity data generated by BT sensors.

Several types of relationships exist within organizations, but, as previously noted, we chose to focus explicitly on task-related and friendship-oriented relationships. These relationships, which have been identified as central elements of informal networks, are particularly relevant to our research objective as they are not stipulated by the organization (e.g., Gibbons, 2004) and, therefore, researchers typically rely on self-reports or third-party observation to capture them. Although these two types of relationships can co-exist, they differ in their nature and central characteristics.

Friendship networks, or networks composed of positive amicable relationships, are social in nature and are characterized by high levels of intimacy, trust, and social support. Friendship is relatively informal and affective in nature, and the interaction patterns that characterize friendships are likely distinct from the interaction patterns that characterize formal, task-oriented relationships. That is, it is plausible that friendship interactions unfold more frequently in constrained spaces (offices, lunch tables, break areas, etc.), and hence, shorter distances.

On the other hand, task-based relationships are more formal. Frequently, these interactions take place in contexts that are less space-constrained, allowing for interactions to occur over greater distances (e.g., conference rooms, small classrooms, etc.). Indeed, prior research suggests that friendship and task-related relationships are differentially related to physical propinquity (Ibarra & Andrews, 1993; Sias & Cahill, 1998). Accordingly, there exists the potential for the differences between primarily friendship and primarily task-related (i.e., advice-oriented) interactions to be reflected in recorded BT RSSI values. Thus, we examine the correlations between BT-generated data at various discrete RSSI values with self-reports of friendship and advice-oriented relationships to address our second research question:

Research Question 2: Can one use discrete Bluetooth RSSI values to discriminate between self-reports of friendship and task-related relationships in a field setting?

Criterion-Related Validity of Bluetooth-Generated Variables for Predicting Performance

Our third research question pertains to the ability of BT-generated data to predict important organizational criteria such as job performance. Indeed, for BT sensors to be considered a valuable supplement to existing methods of data collection they must contribute to researchers' ability to predict outcomes of critical interest in organizational contexts.

When it comes to performing one's job well, an individual's position in the network has been shown to be highly relevant. The literature on interpersonal networks tends to associate different types of benefits with different network structures, but it is generally well-accepted that individuals in more central network positions accrue advantages because they have timelier access to richer and more diverse information. These individuals learn more from their network than their less connected peers, and thus develop greater potential to synthesize and recombine this information into novel ideas (e.g., Burt, 2004; Ebadi & Utterback, 1984; Morrison, 2002).

Research on the relationship between individual (i.e., ego) network structure documents a replicable positive effect of number of brokerage opportunities (i.e., structural holes) on individual knowledge creation (Burt, 2004; Fleming, Mingo, & Chen, 2007; McFadyen, Semadeni, & Cannella, 2009). The positive effect of brokerage opportunities on individual performance should be particularly salient in knowledge-intensive organizations characterized by high task interdependence and extreme levels of specialization (Clement, Shipilov, & Galunic, in press). Accordingly, we expect that in the field context of our first study, which involves people working collaboratively on a large scientific installation, an individual's performance increases with the number of his or her brokerage opportunities. Thus, our third research question:

Research Question 3: Can one use network measures derived from wearable Bluetooth-generated data to predict future job performance in a field setting?

Variance Decomposition of Daily Networks: Within- versus Between-Unit Variance

Although assessing the ability of BT-generated co-location data to (a) display convergent validity with self-reports of relational data, (b) show discriminant validity with respect to qualitative differences in relationships, and (c) provide predictive validity with respect to performance are all important objectives, each of those objectives can be accomplished without fully taking advantage of the unique strengths of this technology. To accomplish the first three objectives, one simply aggregates the BT-generated data to arrive at a single "unit" score for the focal dyad (convergent and discriminant validity) or individual (predictive validity). However, as we stressed at the outset, what makes wearable BT sensors a particularly disruptive technology when used in an organizational context is their ability to collect continuous, objective, highly granular data over extended periods. This creates the opportunity to construct network graphs on a weekly, daily, or even hourly basis – an opportunity that is unprecedented in organizational research.

For example, over the course of a year BT technology could enable one to generate 365 daily network graphs for a particular context. One could "average over" these graphs to assign a single score to each unit (where the unit could be at the individual level, the dyadic level, or the whole-network level) and use this single score to make between-unit comparisons. Indeed, this would be reflective of approaches traditionally taken in network research. However, the failure to recognize the variability in the 365 networks that display themselves throughout the year would be a major missed opportunity. Simply examining the "average" network or individuals' "average" positions within the network glosses over a host of potentially meaningful

configurations based upon time that are worthy of study. These temporally-bounded networks may reflect incremental linear network evolutions, cyclical seasonal variations, or discontinuous re-configurations as the group confronts different tasks or responds to external shocks, all of which are systematic and worthy of theory and empirical examination.

Regrettably, the labor-intensive nature of existing data collection methods for capturing networks not only makes it unfeasible to study such possibilities but also precludes our ability to ask even the most basic question: how much of the variance in traditional network measures is within units and how much is between units? As we noted earlier, research on constructs that were traditionally or solely approached from a between-unit perspective, but then approached from a within-unit perspective, has consistently found that very large portions of variance in constructs reside within units when they are measured over time (e.g., Dalal et al., 2009; Johnson, Lanaj, & Barnes, 2014; Matta et al., 2017; Stewart & Nandkeolyar, 2007; To et al., 2012). Methodological approaches that make this formerly "invisible" variance visible and analyzable create new and innovative opportunities for theory building, theory testing, and exploration. For example, a dynamic, within-unit perspective opens the door for a closer examination of individual agency, which could prove relevant for research on tie stability (i.e., the stability of interpersonal connections) as well as network structural periodicity (i.e., the regularity of activity states; cycles of activity). Thus, our fourth research question:

> *Research Question 4:* Can daily co-location networks derived via Bluetooth data be used to (a) partition the variance associated with network parameters into between- versus within-unit portions, and (b) detect a major discontinuous change in a network's temporal evolution?

Incremental Validity of Bluetooth Data for Capturing Past Events

In addition to predicting future events, BT may provide additional value when it comes to capturing past events. Research has consistently documented human memory limitations associated with retrospectively reporting time spent with various groups and individuals (Freeman, 1992; Marsden, 1990; Schwarz, 1999). Retrospective inaccuracies in self-reports have been attributed to several causes, including the use of simplifying heuristics, social desirability biases, demand characteristics, and self-serving attributions (Collopy, 1996; Gosling, John, Craik, & Robins, 1998). The objective nature of BT signals helps overcome these biases and, when used in conjunction with self-reports, BT reports of time spent with different groups or individuals in the past may provide incremental validity when it comes to recalling a known criterion from the past. Thus, our fifth and final research question:

Research Question 5: Does Bluetooth-generated data provide incremental validity, above and beyond self-reports, with respect to capturing past behavior in a field context?

This last question was prompted by a reviewer comment to a previous version of this manuscript, and because it required a "ground truth" score this question was addressed in a different context than what was used for the first four questions. This allowed us to test another recommendation made by Chaffin et al. (2017) with respect to single sensor versus multi-sensor platforms. Specifically, Chaffin et al. (2017) made a blanket recommendation that the field of WSTs move away from multi-sensor platforms that place more than one sensor on a loosely attached device, and instead focus on "single purpose platforms" that are optimized around one sensor.

There were four reasons for this recommendation. First, it is easier to optimize around one technology than four simultaneously. Second, when optimized around a single sensor (such as BT), battery life for the platform can be extended far beyond what is required to power four

different sensors (especially energy draining microphones). Third, a single purpose BT sensor is far less cumbersome and can be attached in a manner that is far less obtrusive than a platform that needs to capture infrared detections and voice. Finally, the cost of purchasing and processing data from single sensor platforms is far less than what is associated with a multi-sensor platform.

Thus, this second study reports the results from employing a single sensor, low cost, long battery life, wearable BT device. This device was deployed in a context where supervisors of MBA teams observed these teams as part of their formal assignments for the MBA program. This context was heavily covered by cameras and, hence, we were able to obtain ground truth assessments of how much time each supervisor spent with each team and compare this "true score" to both self-reports and BT-generated data. In addition to providing evidence of incremental validity, this allowed us to assess the convergence between two different platforms for capturing BT signals, thus providing a check on the assumption that BT sensors are standardized commodities and that their data streams should be largely independent of "platform" (i.e., manufacturer) attributes.

Study 1

Method

Research site and sample for Research Questions 1 through 4. Our first field study took place in the context of an $800 million, large-scale scientific project aimed at building the next generation linear particle accelerator in the United States: the Facility for Rare Isotope Beams (FRIB). The FRIB is a collaborative effort between the United States Department of Energy Office of Science (DOE-SC) and a large Midwestern university, and is a critical component for scientific initiatives that seek to understand the fundamental forces and particles of nature as manifested in nuclear matter. Scientific research from FRIB will enable the nuclear

science community to make discoveries about the properties of rare isotopes (short-lived nuclei not found on Earth) and major advances in the understanding of nature by producing isotopes that previously only existed in the most violent conditions in the universe (York et al., 2009).

At the time our study was taking place, the linear accelerator was being designed and built by engineers who worked closely with the nuclear scientists who would then use the accelerator for research on rare isotopes. The engineers and scientists were supported by administrative assistants as well as an over-arching leadership group that coordinated the efforts of the engineering, scientific, and administrative employees. Our focus was on a large, multifunctional team that was comprised of 32 of the top representatives of three functional areas (i.e., Engineering, Scientific, and Business Support) and Leadership. Though all individuals were part of one large team, for ease of exposition we will refer to the groups of functionally distinct employees as "teams" and the larger team, as a whole, as a "multiteam system."

The study took place over four weeks and included one major discontinuous event, the groundbreaking at the facility, which was roughly symbolic of the shift from the design stage to the construction stage. Typically, the United States Department of Energy commissions a new accelerator every 20 to 25 years. Due to changes in technology over that period, along with the unique focus for each new accelerator, this means that, in the words of FRIB leadership, "This specific accelerator had never been built before and would never be built again." Thus, this was a knowledge-intensive organization where it was difficult to fully specify in advance exactly who needed to work with whom as the process unfolded. FRIB draws upon the interdisciplinary collaborative efforts of individuals from a wide variety of occupational specialties. Formal job titles range from generic jobs typically found in almost all organizations, such as "Human

Resource Administrator," "Financial Analyst," and "Clerical Assistant," to specialized jobs, such as "Rare Isotope Beam Physicist," "Dipole Fabrication Specialist," and "Cryogenic Engineer."

In this study, we used the WST platforms offered by Humanyze (formerly Sociometric Solutions), which was also the platform used in the controlled laboratory studies reported by Chaffin et al. (2017). At FRIB, we collected data on different days over a four-week period. Our data collection resulted in 9 WST deployment days that we numbered -6 to +18, reflecting the number of days prior to and after the groundbreaking event. The observation period per participant varied slightly over this time based on on-site attendance patterns (i.e., different work habits and off-site assignments). However, in all cases, we considered the observation period as sufficient to allow true on-site interaction patterns to emerge. Given that these BT sensors perform multiple scans for other BT sensors per minute, the number of observations we obtained across the 9 WST deployment days was well into the hundreds of thousands.

Bluetooth detections. Over the length of 9 days, the wearable sensors attached to the 32 individuals who were top-level representatives of the Engineering, Scientific, Business Support, and Leadership Teams detected a total of 220,192 co-location events. We applied a two-step procedure to purge spurious detections from our data. First, we needed to eliminate proximity detections that were not associated with human interaction opportunities. Thus, we restricted our sample to only include records within an RSSI range of -69 to -91 (with an RSSI of -80 in the center). We chose this range as an RSSI of -80 has been identified as an appropriate threshold for face-to-face interactions by prior researchers (Finnerty et al., 2014; Olguin et al., 2009), and because the pattern of correlations in the data we collected suggested that the strongest correlations between self-reported variables and RSSI detections were in this range.

We interpreted records with RSSI > -69 as spurious close-range detections produced in pre- or post-deployment situations after devices were jointly placed on collection carts. We considered records at RSSI < -91 to reflect spurious long-range detections indicating a distance of separation that would not constitute an opportunity for human interaction. This restriction reduced our dataset to 205,860 records. Second, we manually searched for and removed records that exhibited improbable detections for the 992 dyads (32 x 31) to enhance data integrity. These detections were those that were generated after-hours, when the BT sensors were in the possession of the research team, likely due to the failure to turn the sensors off upon collection. Ultimately, this resulted in the removal of an additional 3,762 detections (1.7% of total data collected), providing us with a final sample of 202,098 detection records.

Figure 1 illustrates the temporal variation of detections across the system. The cumulative number of daily Bluetooth detections range from a minimum of 6,807 on Day -6 (i.e., 6 days before groundbreaking) to a maximum of 51,863 on Day +16 (i.e., 16 days after groundbreaking). Figure 2 describes how individuals differ in their proximity to others. The cumulative number of individual Bluetooth detections range from a minimum of 452 to a maximum of 12,556.

Figures 3a through 3j show the cumulative and daily network graphs where blue dots reflect Engineers, green dots Scientists, black dots Business Support, and red dots Leadership. Figure 3a connects dyads that have 60 or more BT detections over the course of the 9 days within our observed RSSI range (-69 to -91) based on a force-directed graph drawing algorithm (Kamada & Kawai, 1989). We chose 60 BT detections because it represented the median of aggregate detections across all dyads, maximizing the variance in terms of tie presence and absence. Figures 3b-3j connects dyads that have 7 or more BT detections on that given day using

the same algorithm. We chose 7 BT detections because the aggregate average count was produced by 9 daily observations ($60/9 = 6.67$). It is important to note that this agnostic algorithm was able to largely group individuals into their respective functional backgrounds (e.g., Engineering), such that the physical proximity reflected in the graphs was emergent, and not specified in any formal way (i.e., wearable BT data allowed us to detect "known functional groups"). At this point, we provide this data merely to show the structure of the system studied, and how this varied within and between units over time.

     Measures. Three months prior to the WST deployment, we administered a sociometric survey to obtain self-report data about the nature and strength of relationships among participants. We chose to collect these measures prior to WST deployment, rather than concurrently with WST deployment, not only to reduce participate fatigue but also because participants typically report on long-range, enduring patterns of interaction, rather than specific instances, when asked to report on their relations with others (Freeman, 1992; Marsden, 1990). Furthermore, asking participants to report exclusively on their *current* pattern of interactions is apt to be highly problematic as it is often difficult for participants to report on interactions that take place within highly specific timeframes (Bernard, Killworth, & Sailer, 1981). Thus, we asked participants to report on their typical social relations with others.

     We collected this data following procedures similar to those outlined by Reagans, Zuckerman, and McEvily (2004). Due to the large size of the formal organizational network, we used a hybrid fixed roster and free recall method (Wasserman & Faust, 1994). Each respondent was first presented with a list of all FRIB departments and asked to indicate whether they interacted with individuals from each department. Next, we presented a fixed roster of all individuals formally assigned to each department, based on the organizational chart. Finally, we

provided a free response section for up to 15 additional individuals that the participant could write-in.

Thus, our procedures complemented one another such that (a) the fixed roster supplements the free recall questions by ensuring that less salient workplace relationships were reported on, and (b) the free-recall questions supplement the fixed roster by ensuring that important relationships not captured by the fixed roster were included. As noted, participants reported on friendship and task-related relationships, specifically. Presence and strength were captured for task-related relationships using a five-point, Likert-type scale where we distinguished between task-related relationships that reflected "advice-giving" versus "advice-receiving." The presence of friendship relations was captured using a dichotomous variable, which is standard practice in the network literature.

Finally, six months after WST deployment we obtained performance ratings for participants from their supervisors. Specifically, job performance was measured using eight-items adapted to our context from Tsui, Pearce, Porter, and Tripoli (1997). This is a standardized and widely used measure of performance that was specifically designed to be used in contexts where individuals are occupying varied roles. Sample items included the stem "This person", followed by "is very efficient," "strives for higher quality work than is required," "displays effective judgment," and so on. Supervisors rated their subordinates on a five-point scale (1 = *strongly disagree*, 5 = *strongly agree*). Performance ratings were obtained for 31 of the 32 focal participants (96.8%), where the lone missing person was the director of the facility who was not formally evaluated by anyone that was accessible by the research team. The internal consistency estimate of reliability for this scale as operationalized by coefficient alpha was .90.

Analyses. Analyses were conducted in Stata 13.1 (StataCorp, 2013), UCInet 6 (Borgatti, Everett, & Freeman, 2002), R statnet (Handcock, Hunter, Butts, Goodreau, & Morris, 2008), and Mplus 6.12 (Muthén & Muthén, 1998-2011). In approaching our first two Research Questions, where we focused on analyses at the dyadic level of analysis ($n = 992$), we examined BT detections in two ways.

Analytic approach to Research Question 1. To address the first research question, convergent validity, we tested the degree to which 992 self-reports of friendship relationships and two kinds of task-related relationships, advice-giving and advice-receiving, correlated with increasingly liberal ranges of the WST wearer's *cumulative* BT detections. Cumulative BT detections represent the sum of all detections generated at distances up to a particular RSSI "threshold." For example, at a cumulative RSSI threshold of -80 we are including and examining all detections generated at RSSI values (i.e., detection distances) greater than or equal to -80, such as -79, -78, -77, and so on, as less negative RSSI values indicate closer distances (i.e., an RSSI of -79 indicates a shorter distance than an RSSI of -80, while an RSSI of -81 indicates a larger distance). This is how BT detections are typically examined in WST literature. We did this not only to examine the convergence between self-reported relational variables and BT detections, but also to illustrate the effects of different thresholds on the correlation between cumulative BT detections and self-reported relational variables, as thresholding represents a widespread practice in the burgeoning literature on WSTs.

When one engages in thresholding, one essentially removes all BT detections with RSSI values that are less than a particular value. For example, by using a threshold of RSSI = -80, one would remove all BT detections generated at RSSIs of -81, -82, -83, and so on, with the assumption that BT detections registered at these RSSI values indicate distances too great to

represent a legitimate interaction. Though the logic underlying thresholding may make sense to an engineer, we argue, from a psychometric perspective, that this practice restricts correlations and limits one's ability to predict task-based (i.e., advice) and affective (i.e., friendship) relationships.

Thus, in sum, the correlation between cumulative BT detections and sociometric self-reports serves as an indicator of convergent validity. The rationale behind using cumulative BT detections is to illustrate the effect of choosing increasingly inclusive, dichotomous RSSI thresholds (i.e., greater spatial separation), bearing in mind that some researchers have limited themselves to fairly restrictive thresholds (e.g., Onnela et al., 2014). In doing so we explore the extent to which the use of RSSI thresholds may affect the potential utility of BT sensors. This is of relevance as BT sensors generate detections at a larger range compared to alternative WSTs based on infrared technology (Chaffin et al., 2017).

Analytic approach to Research Question 2. To assess the potential of BT data to detect qualitative differences in interactions, we shifted our analytical focus from *cumulative* RSSI thresholds to *discrete* RSSI levels. This allows us to discriminate between alternative spatial-separation situations. That is, every distinct RSSI value was considered as a separate category or "bin" with the corresponding number of detections. When we look at discrete RSSI levels, we are specifically examining detections registered solely at a particular RSSI value. For example, when we correlate self-reports of friendship with BT detections generated at an RSSI of -80, we are including only those BT detections generated at that particular RSSI value (which is indicative of a particular distance). Thus, detections generated at RSSI values greater than (e.g., -78, -79) or less than (e.g., -81, -82) -80 are not included in this calculation.

The purpose of examining discrete RSSIs is to examine the possibility that BT detections registered at different RSSI values capture different relationship types (i.e., discriminant validity). For example, friends might be expected to stand or sit closer together than colleagues, a pattern that could be reflected in a relatively stronger relationship with a particular RSSI value corresponding to the pair's preferred interaction distance. Evidence that BT RSSI can be used to discriminate between alternative relationship types would add value to BT sensors as a major limitation associated with WSTs is that they lack contextual information (e.g., does this interaction represent an affective or task-oriented interaction?). Analogous to Research Question #1, we calculated the correlations between BT and self-report data.

Recent empirical research on BT proximity sensor data suggests that restricting analyses to close proximity detections improves the correct identification of particular interactions, such as friendship interactions (Sekara & Lehmann, 2014). Thus, to better understand when BT proximity data are particularly effective in predicting certain relationships, we also employed exponential random graph modeling (ERGM) techniques. ERGMs express the probability of a given whole network structure with respect to a set of parameters, and are analytically appropriate for modeling the factors driving interdependent dyadic relationships within a social context relative to more traditional estimation approaches that rely on non-interdependence assumptions (Snijders, Pattison, Robins, & Handcock, 2006). In our model specifications, we consider BT detections as a dyad-level exogenous covariate. We then evaluate whether thresholding improves or degrades model fit when it comes to predicting actual self-reported friendship, advice-giving, and advice-receiving networks in a "low information context" (i.e., no knowledge of functional background, such as Engineering, Leadership, and so on). We then re-examine this same question in a "high information context" that includes additional variables

that can be calculated when one has information on functional background. Rather than focus solely on friendship networks (Sekara & Lehmann, 2014), we also examined advice networks to see if different conclusions regarding the value of thresholding could be drawn based upon the network in question.

Analytic approach to Research Question 3. When examining the criterion-related validity of BT data with respect to individual performance, we shifted the level of analysis from the dyad level to the individual level, where we simply regressed supervisory ratings of performance on the number of BT data-generated brokerage opportunities, controlling for aggregated, self-reported friendship, advice-giving, and advice-receiving.

Analytic approach to Research Question 4. To decompose variance into within- and between-unit components, we measured two widely used parameters within the network literature and scored individuals on these metrics. These parameters include ego network density and heterogeneity. Ego network density reflects the portion of possible ties among an individual's direct contacts that are realized and is inversely related to an individual's brokerage opportunities. On the other hand, heterogeneity is based on the count and proportions of "alter" memberships in different functions and reflects the diversity of knowledge resources an individual can access. We calculated Blau, Blum, and Schwartz's (1982) measure of heterogeneity as implemented in UCInet (Borgatti et al. 2002).

First, we aggregated across the 9 days the study took place and scored individuals via a purely "between units" approach. This would be traditional practice with self-reported network data and one could use this data to infer that a person has greater access to diverse knowledge than another person (i.e., heterogeneity) or that a person has fewer brokerage opportunities than another person (i.e., density), in general. We then treated each daily observation as a meaningful

unit and scored individuals in a daily fashion. This allowed us to assess variability within units and created the opportunity to see if some individuals' ego networks varied daily (e.g., if individuals are in contact with diverse knowledge on some days but not others).

If all the variability existed between units, each person would have the exact same score every day, meaning that there would be *zero within-unit variance* across days. In other words, all the graphs depicted in Figures 3a through 3j would be identical. On the other hand, if all the variability existed within units, all respondents would have the exact same average across the nine days *but would vary widely* from day to day. Neither of these extremes is likely and, thus, this total variance can be decomposed into within and between components that sum to 1.00.

To establish if variance within units is indeed systematic and not simply random error, we then factor analyzed the dyadic network data to see if the interaction patterns were systematic and related to known changes in the nature of the work conducted in the organization. More specifically, we wanted to see if the factors that emerged were representative of distinct activity states before and after the official groundbreaking ceremony. In this factor analysis, the rows reflect dichotomous (0,1) determinations as to whether any 2 people were linked on the day in question, and the columns reflect the 9 days where people could have possibly been linked. Thus, if Person #1 was linked to Person #2 on Day -6, this value would be 1 for that pair on that day. If Person #1 was not linked to Person #3 on that day, that value would be 0. Thus, the matrix that was factor analyzed contained 992 rows, 1 for each dyad, and 9 columns, 1 for each day.

If the variance within dyads was due to random error, then no strong factors would emerge from this analysis. However, if there were systematically recurring interactions (e.g., Person #1 was linked to Person #2 most of the days, but rarely was linked to Person #3) these recurring interactions would be reflected in a single over-arching factor that would capture this

systematic variance. That is, if the pattern of interactions across dyads was systematic and stable across time, then one general factor would emerge from this "Dyad x Day" matrix that would reflect, in general, whom was systematically linked to whom. However, if systematic interaction patterns varied by day, such as before or after the groundbreaking, then this would mean that there were systematic interaction patterns that varied over time. So, for example, Person #1 may have had one set of systematic interaction patterns *before* the groundbreaking, but then had a different, but still systematic, set of interaction patterns *after* the groundbreaking.

Results

Evidence regarding convergent validity. The correlations between *cumulative* BT detections at increasingly inclusive RSSI thresholds and self-reported relational data are presented in Table 1 (along with the average standard error for those correlations). As a benchmark by which to judge the estimates of convergent validity between BT detection data and psychological constructs, Meyer et al. (2001) compiled meta-analytic estimates of cross-method convergent associations for a wide range of psychological variables. Generally, correlations between different methods were moderate in size, such as correlations between supervisor-report and peer-report ratings of job performance ($r = 0.34$), and correlations between self-report and peer-report ratings of job performance ($r = 0.19$). Moreover, Eagle et al. (2009) compared self-reports of proximity with average daily proximity based on BT scans and found that correlations between observed (i.e., BT) and reported (i.e., survey) proximity among individuals ranged from $r = .16$ to $r = .41$.

In the context of our study, we find that correlations between cumulative RSSI thresholds and reported friendship ties increase to a maximum value of $r = 0.34$ ($p < .01$) at an RSSI of -83. The correlations between cumulative RSSI thresholds and reported advice-giving relationships

reach a maximum of $r = 0.48$ ($p < .01$) at an RSSI of -83. For advice-receiving relationships, the

strongest correlation with cumulative RSSI is $r = 0.53$ ($p < .01$) at an RSSI of -89. Thus, in

general, the answer to Research Question #1 seems to be that BT detection data do converge

with self-reports of friendship and task-related relationships, and that the highest level of

convergence may occur beyond the commonly used RSSI threshold of -80. The importance of

the nature of the relationship assessed and threshold value also must be recognized, in that the

amount of variance explained ranges from 6% to 28% depending on these distinctions.

Evidence regarding discriminant validity. The correlations (and their average standard

errors, see table *Note*) between BT detections at *discrete* RSSI levels and self-reported relational

data are presented in Table 2. Importantly, quadratic assignment procedure (QAP) correlations

(5,000 permutations) among the self-reported friendship, advice-giving, and advice-receiving

networks suggest that the two advice networks were highly correlated with one another ($r = .80$,

$p < .05$) whereas the friendship network was moderately correlated with both the advice-giving

($r = .43$, $p < .05$) and advice-receiving ($r = .37$, $p < .05$) networks. This suggests that individuals

who give advice to each other often receive advice within the same relationship. However,

though correlated, individuals who have advice-oriented (i.e., task-oriented) relationships with

one another may not necessarily be friends. Therefore, advice and friendship networks likely

represent unique networks, which may manifest themselves differently when it comes to BT

detection patterns (i.e., RSSI values).

Indeed, we observe the highest correlation between RSSI levels and reported friendship

($r = 0.35$, $p < .01$) at the RSSI value of -76. After this, the correlations decline steadily. For

advice-giving, correlations increase to a maximum ($r = 0.46$, $p < .01$) at the RSSI of -83, and

then decline. This pattern is repeated for advice-receiving where the maximum correlation ($r =$

0.51, $p < .01$) is calculated at an RSSI of -82. Notably, the peak correlation for friendship is noticeably smaller than the peak correlations obtained for advice-giving and advice-receiving. This is attributable to the dichotomous scale used to measure the friendship variable and the ordinal scales used to measure the advice-giving and advice-receiving variables. The format of the scales of the advice variables allow for greater variance, and higher variance facilitates larger correlations among variables (Cohen, Cohen, West, & Aiken, 2003).

Hence, range restriction in the friendship variable may account for the lower peak correlation. To illustrate what the correlation between friendship and RSSI would have been had friendship been measured with an ordinal scale, we calculated a biserial correlation (not a *point* biserial correlation). Biserial correlations are correlations in which one of the variables is a dichotomized variable that might have underlying continuity. For example, friendship could have been measured on a Likert-type scale ("How strong is your friendship with each of the following individuals", on a 1-5 scale), but can also be classified as friends/not friends. To calculate the biserial correlation, we used the formula in Cohen et al. (2003):

$$r_b = ([M_{Yp} - M_{Yq}]*PQ)/h(sd_Y)$$

Where M_{Yp} and M_{Yq} are the means for the two points of the dichotomy, P and Q (= $1 - P$) are the proportions of the sample at these two points, and h is the ordinate (height) of the standard unit normal curve at the point at which its area is divided into P and Q portions.

This calculation yielded a biserial correlation of $r_b = .54$ between our friendship variable and BT detections at a discrete RSSI value of -76, or the peak friendship-RSSI correlation identified in Table 2. This correlation is comparable to the peak correlations observed for advice-giving ($r = .46$) and advice-receiving ($r = .51$) in this same table. Thus, in terms of Research Question #2, our results support the idea that BT RSSI information may be leveraged as a filter

for identifying proximity events more likely to be associated with a particular interaction type.

We see the strongest convergence between BT detections and reported friendship at RSSI levels

that suggest close proximity situations. In contrast, convergence of BT detections with reported

advice-receiving and giving relationships was highest at more liberal RSSIs, indicating greater

spatial separation. Again, the importance of the nature of the relationship assessed and threshold

value must be recognized, in that the amount of variance explained ranges from 6% to 26%

depending on these distinctions.

Dichotomized thresholding: ERGMs. As we noted earlier, in engineering contexts

thresholding is viewed as a data reduction technique where the value of simplifying the data

structure by dichotomizing all the values outweighs the value of keeping the data in a continuous

form. This is anathema to practice in psychometric contexts where it is considered empirically

wasteful to convert continuous data into dichotomies. Because WST development has, to date,

been driven by engineers, the practice of thresholding is widely adopted in this field without

question. As part of this construct validation work related to convergence and divergence, we

attempted to bridge this engineering-psychometric gap by evaluating the potential liabilities and

benefits of BT RSSI thresholding. Specifically, we compared two alternative ERGM

specifications for modeling the self-reported friendship and advice networks, integrating findings

from our prior analyses related to convergence and divergence.

Based on extant theory we first built ERGMs that included three basic effects labeled

"Density," "Reciprocity," and "Daily Average BT Detections." The "Density" effect accounts

for the network level probability of friendship relations relative to a random graph where the

population rate for friendship is assumed to be .50 (i.e., half the people within the population are

characterized as friends). For example, a statistically significant negative coefficient for

"Density" implies a "sparse" friendship network (i.e., fewer actual ties than expected by chance if half the people were friends). The "Reciprocity" effect describes the propensity for individuals to identify each other as friends. For example, a statistically significant positive "Reciprocity" coefficient implies that one is more likely to report friendship ties if he or she is frequently identified as being a friend by others. The "Daily Average BT Detections" coefficient implies that one is more likely to report friendship ties if he or she is often detected as being collocated with others based upon BT detections. We tested this base model with an unrestricted (without thresholding) and restricted (with thresholding) BT detection count variable as a dyadic covariate to see whether thresholding degraded the fit of the model.

The ERGMs tested in Tables 3, 5, and 7 presume no information regarding the formal organizational structure that describes the multiteam system in which these individuals were embedded (i.e., functional background), and hence, might be considered "low information" contexts. For Tables 4, 6, and 8, we changed this and specified ERGMs that included the variables in Tables 3, 5, and 7 as well as a new set of variables that reflected structural knowledge regarding functional background. Specifically, the "Functional Homophily" effect reflects the fact that individuals tend to form friendships or advice-based relationships with similar others, which in this case means being part of the same functional team. For example, a statistically positive coefficient for "Functional Homophily" indicates elevated odds of friendship ties or advice-oriented ties occurring between individuals that share the same functional classification.

We then added three dummy variables that contrasted members of specific functional teams with each other, where the Engineering Team was the contrast for the Business Support Team, Scientific Team, and Leadership Team (i.e., the Engineering Team was coded with all

zeros). An effect here indicates elevated odds of friendship or advice ties for people in specific

teams when compared to the Engineering Team. Thus, a significant and positive effect for the

Science Team means that people who are part of this team are more likely to have friends, or are

more likely to give or receive advice, compared to people in the Engineering Team. As in Tables

3, 5, and 7, we then expanded the base model with an unrestricted and restricted BT detection

count variable as a dyadic covariate to see if thresholding (at BT RSSI > -80) improved or

degraded model fit relative to an unrestricted detection count.

    First turning to Table 3, in the "low information" context, all three variables were

statistically significant in both models. However, the fit for the model that did not employ

thresholding was better (lower scores reflect better fit for both Akaike and Bayesian indices, AIC

and BIC) relative to the model that did use thresholding. The same can be said for the other "low

context" ERGMs, the results for which are in Table 5 (advice-giving) and Table 7 (advice-

receiving). Thus, thresholding did not enhance our predictive ability, no matter the dependent

variable, in "low information" contexts.

     Turning next to Tables 4, 6, and 8, or the "high information" contexts, all three

previously significant effects were also significant here. In addition, there were significant

effects for "Functional Homophily," as well as for the one dummy variable that represented the

Scientific Team versus Engineering Team contrast in the friendship ERGMs and the one dummy

variable that represented the Leadership Team versus Engineering Team contrast in the advice

ERGMs. Not surprisingly, regardless whether one employed thresholding, the fit for the models

reported in Tables 4, 6, and 8, which incorporated information on functional background,

provided better fits to the data relative to the models tested in Tables 3, 5, and 7, which did not

incorporate information on functional background. Most importantly, however, the fits of the

ERGMs in Tables 4, 6, and 8 that did not use thresholding were superior to the fits of the

ERGMs that did employ thresholding. Thus, in both low and high information contexts, and

across all three dependent variables (i.e., friendship, advice-giving, and advice-receiving), fit was

better when the data was left in its continuous form relative to when one used the recommended

threshold value for dichotomizing the BT data.

Thus, the common practice of including only those BT detections with RSSI values

greater than -80 (e.g., -79, -78, -77) does not enhance our ability to predict directed friendship,

advice-giving, or advice-receiving, regardless of whether we have knowledge of the formal

organizational structure (Tables 4, 6, and 8) or not (Tables 3, 5, and 7). This suggests that the

practice of thresholding, widely used in this emerging literature as a means of data reduction, is

not only unnecessary but may also degrade one's ability to predict outcomes of interest.

Evidence regarding criterion-related validity. Criterion-related validity reflects an

instrument's ability to predict future outcomes (Rosenthal & Rosnow, 2008) and, therefore, one

could argue that BT sensors represent valid, valuable measurement instruments if the data they

generate predict outcomes of interest. Given the knowledge-intensive nature of the host

organization, we expected to find that the more an individual occupied a position of brokerage

between other individuals, the higher would be his or her rated performance. A broker in a

network fills structural holes, connecting otherwise disconnected parties and serving as a

mediator or gatekeeper of information between them (Burt, 1992). These individuals learn more

from their network than their less connected peers and thus develop greater potential to

synthesize and recombine this information into novel ideas (e.g., Burt, 2004; Ebadi & Utterback,

1984; Morrison, 2002). Given the robust and well-accepted nature of this relationship, we

decided that it represented a suitable variable for testing criterion-related validity.

Our brokerage position variable was created using the BT-generated, aggregated count data and UCInet 6 (Borgatti et al., 2002), and it is indicative of the number of times a participant resides on the shortest structural path between two other individuals. Table 9 provides the correlations among variables at the individual level of analysis. Friendship was aggregated as a sum of self-reported friendship ties given the dichotomous nature of the variable while advice-giving and advice-receiving were aggregated as averages given the ordinal scales used to capture these variables. In relation to Research Question #3, brokerage opportunities predicted future job performance ratings obtained from direct supervisors (β = .44, p < .05), controlling for friendship (β = .10, *n.s.*), advice-giving (β = .15, *n.s.*), and advice-receiving (β = -.38, p < .10), when all variables were included in a multiple regression. These results suggest that BT-generated brokerage is indeed predictive of performance.

Evidence regarding within-unit and between-unit variance. Figures 3b through 3j, shown previously to provide a depiction of the structure of the multiteam system that was studied here, makes two points very evident. First, even though the method used to depict this structure was agnostic to functional background, the BT data largely grouped the 32 people into their respective functions based upon co-location and proximity, thus capturing "known groups" (Nunnally, 1978). Second, there was clear variability in the daily network patterns within and between groups. To provide a meaningful evaluation of how much variation is lost by aggregation in terms that would be appreciated by network researchers, we took two widely used ego network indices (i.e., ego network density and heterogeneity) and calculated the percentage of within- and between-unit variance for each.

With respect to ego network density, a full 82% of the variance resided within individuals and only 18% resided between individuals. This means that to better understand the role of ego

network density, researchers need to recognize that any one person's network density varies depending upon the activity pattern that captures the work being done on a given day. The same is true with respect to ego network heterogeneity, where 62% of the variance is within individuals and 38% of the variance is between individuals. Thus, when it comes to conceptualizing ego network heterogeneity, researchers have been ignoring the extent to which any one person's network heterogeneity is contingent upon the work configuration associated with different days. A great deal of variance is lost by aggregating over days and just asking: "on average, what does this person's network look like in terms of density and heterogeneity?" We cannot emphasize enough that no prior study has been able to study "daily networks" like this, and, hence, all extant research may be missing where the majority of variance lies in network parameters.

Although there is a great deal of daily variance in network parameters that is lost when one averages across days, one may be concerned that this variance is random rather than systematic. To assess whether this data was indeed systematic, we factor analyzed all the potential dyadic connections that might exist between each possible pair of individuals and treated those as "units" ($n = 992$) where each entry was either 0 or 1 (analogous to whether a person got an item right or wrong on a standardized test). We then treated days as "items" so that a factor analysis of this matrix might reveal systematic collaborative interaction patterns when days load together on the same factor (i.e., pre- and post-groundbreaking).

If the daily data is random, then no strong factors will emerge. If each day has the exact same collaboration pattern, then one single over-arching factor will emerge. If each day is unique, and is associated with its own idiosyncratic collaborative pattern, then there would be a separate factor for each day. Finally, if some days are more alike than other days, which are in

turn, more alike than other days, then some small set of factors would emerge. Because our data collection period bridged the groundbreaking event when the FRIB project "officially" shifted from design to construction, we expected to find two factors that captured the work patterns prior to and after the groundbreaking. Because we were not truly subject matter experts of FRIB operations, however, we used an exploratory analysis to allow for the emergence of unanticipated factors.

Factor analytic results with varimax rotation are shown in Table 10, which demonstrates that two factors emerged from this analysis. We chose varimax rotation as we expected distinct activity states to emerge, and we settled upon a two-factor solution for several reasons. First, a "scree test" (Fabrigar, Wegener, MacCallum, & Strahan, 1999), used to visually examine eigenvalues, indicated that the last substantial drop occurred with the second factor. Second, a three-factor solution provided unacceptable results as the factor loading for Day +16 was greater than 1.0 and was paired with a negative residual variance, suggesting over-extraction. Finally, given the known change in activity states that occurred during the data collection period, a two-factor solution made greatest theoretical sense. Thus, the two-factor solution was retained.

The first factor clearly captured the days prior to the groundbreaking event and reflected strong collaboration patterns within the Business Support Team and between the individuals working within the Business Support Team and individuals on all the other teams, particularly Engineering and Leadership. The second factor captured the days after the groundbreaking event and reflected very strong collaboration patterns within the Engineering Team as well as interactions involving individuals on the Engineering Team and other teams. Their eigenvalues were 3.025 and 1.477, respectively. Consultation with subject matter experts (SMEs) at FRIB suggested that these two factors correspond to two different types of "operational modes" that

are characteristic of the center. Our SMEs suggested that the first factor reflected a "public site visitation" mode where preparation for external visiting dignitaries was being conducted. As an $800 million publicly-funded, next generation particle accelerator, the FRIB was often a target of inspections or visits from funding agencies (DOE or NSF), politicians (national and regional), scientists (from other national or international facilities), government agencies (EPA or Homeland Security), and, in some cases (e.g., the groundbreaking), all the above.

The second factor, according to our SMEs, reflected "normal operational mode," which was driven by engineers at that time because FRIB was in the design and construction process. This literally involved architectural planning, as well as physical digging, pipe installation, and other assorted engineering activities. This factor was most representative in terms of capturing the largest number of days. Taken together, relative to Research Question #4, the evidence suggests that (a) there is a great deal of variance associated with daily networks that has gone unrecognized by prior research, (b) this variance is systematic, and (c) the failure to assess and analyze such variance may be a severe limitation when it comes to understanding the precise and nuanced nature of collaborative patterns as they unfold over time.

Study 2

Although the BT-generated data in the first study demonstrated convergent validity with self-reported relational data (e.g., friendship), the sole reliance on self-reports may raise concerns given the aforementioned issues often associated with them. That is, although the laboratory studies reported by Chaffin et al. (2017) employed known true scores, this was not the case in any of the analyses used to address the first four of our research questions. Furthermore, the value of BT sensors would be enhanced if empirical evidence suggested that they could indeed capture interactions that self-reports would miss. Thus, we conducted a second study in a heavily

monitored field setting. In this study, we examined the extent to which wearable BT sensors could provide incremental validity when it comes to recalling the amount of time team supervisors spent interacting with different teams for whom they were responsible in a context where we had known scores, captured by video (i.e., ground truth).

Method

Research site and sample for Research Question 5. Our second study took place in a laboratory suite on the campus of a large Midwestern university where the research participants worked as team supervisors. The laboratory suite consisted of four rooms, Rooms A, B, and C, and a main room that connected the three lettered rooms. Figure 4 provides a visualization of the suite's physical layout. Three teams at a time worked on a team-based simulation in one of the three lettered rooms, while the main room was used for introductions, debriefing, and interactions among supervisors during the simulation.

Once per week, over the course of 3 weeks, 15 teams came to the lab and completed 3 simulations. Given the capacity of the lab (i.e., 3 teams at a time) and the number of iterations that each team engaged in (i.e., 3 iterations each), there were a total of 45 simulations run over the course of 15 lab sessions (i.e., 5 sessions per week over the 3-week period, 3 teams per session). During each session, teams were observed by various team supervisors who worked for the MBA Program. In total, 9 supervisors observed, interacted with, and engaged in instructional material with these teams. Some supervisors came to as few as a single session, whereas others came to as many as 6 sessions. Collectively, these 9 supervisors attended 26 sessions as more than one supervisor could observe the simulation at once.

Each time supervisors attended a session, they were equipped with a wearable BT sensor. Furthermore, additional BT sensors were planted throughout the suite in the locations that

supervisors tended to congregate: just inside the doors of the four rooms. These planted BT

sensors served as "base stations," and were used to determine the amount of time that supervisors

spent in each of the four rooms (i.e., with each team). Given that the 9 supervisors collectively

attended 26 sessions, and that each session involved teams that the supervisor might interact with

(the three MBA teams and the Instructional Team) in various locations (i.e., Rooms A, B, C, or

the main room), we had a total of 104 (26 x 4) possible person-team dyads. However, due to

software malfunctions, our final sample consisted of 84 person-team dyads (80.77%).

Each session was videotaped and independently coded by one graduate student and one

undergraduate student. Coders kept a record of the number of minutes that each of the 9

supervisors met with each of the four teams (i.e., the three MBA teams and the Instructional

Team), resulting in a running total for each person-team dyad. Agreement among video coders

was extremely high, with an ICC(1) of .97 and an ICC(2) of .98, far surpassing acceptable levels

(Bliese, 2000; James, 1982). Thus, we averaged the two coders' ratings and treated this as

ground truth to be predicted with BT-generated data and supervisors' self-reports.

Bluetooth sensors. The BT sensor data used in Study 1 was captured using the

Humanyze device, a multi-sensor platform. Following the recommendation of Chaffin et al.

(2017), we migrated from the multi-sensor platform used in Study 1 to a single-sensor BT device

offered by Limefy. This device uses less energy to capture BT signals and is not configured as

adjacent to any other sensors (such as microphones or accelerometers), as is the case with the

Humanyze platform.

To ensure that the two WSTs were comparable in their ability to measure variability in

distance based on BT signal strength, we conducted a pre-validation experiment in a barrier-free

environment, varying the distance between each device (ranging from approximately 1 meter to

8 meters). The position of each device was marked so that the placement was consistent across sensor platforms and multiple iterations. Each device was then left in a static position for a period of 45 minutes in our experimental setting. It is important to note that the correlation between BT RSSI and increasing distances should be negative because RSSI should decrease (i.e., become more negative) as distance increases (i.e., becomes larger).

The correlation between actual distance and the RSSI values recorded by (a) the Humanyze BT sensors was $r = -.55$ ($p < .01$) and (b) the Limefy BT sensors was $r = -.72$ ($p < .01$). These results demonstrate the validity of BT signal strength detections in terms of differentiating distances between sensor locations. Additionally, the relationship between the signal strength of the two devices (based on the location each sensor was placed) was $r = .67$ ($p < .01$), providing evidence of convergence between the two devices. Given that the BT sensors on the two devices differentiated distance and correlated highly with one another, we migrated to the single-sensor Limefy platform.

Measures. Two weeks after these sessions took place, the 9 supervisors reported the amount of time that they had spent interacting with various teams in the four rooms of the suite during the sessions they attended. To facilitate the supervisors' memories, they were given a very precise amount of time to allocate among the rooms/teams, such that the amount of time they allocated corresponded to the number of minutes captured by the video camera (and consequently coded by coders). Furthermore, they were given specific information about which teams were in which rooms (as each supervisor had primary responsibility for specific teams), the time at which labs began and ended, and the dates of each session, along with a diagram of the room. Thus, these conditions were quite favorable for the supervisors in terms of recall, and perhaps much more structured relative to what one might see in a less structured organizational

context, hence providing an upper bound for the human ratings. In contrast, the small space of

the suite and short distance between BT base stations made this a challenging context for the BT

sensors.

   Analytic approach. As noted, BT sensors were placed in the four rooms in locations

where supervisors tended to congregate. These sensors served as base stations and swept for

other BT sensors every 10 seconds. Thus, they were what we used to infer that supervisors were

with various teams. Furthermore, these base stations were used to determine the RSSI level at

which detections were not meaningful. Specifically, we used the average RSSI at which these

base stations detected one another (i.e., detected BT sensors in other rooms) to determine the

RSSI at which we removed detections. Although we do not endorse threshold-based logic when

using BT sensors to capture interactions among individuals, we felt it was appropriate in this

context given the small size of the suite, how closely the base stations were situated to one

another, the static positioning of the base stations, and the fact that we were attempting to

determine time spent in specific locations (rather than detect human interactions).

Results

   Detections generated. 67,508 BT detections were generated among the 4 base stations

and 9 supervisors over the 3-week, 15-session period. Of these, 55,915 were between the 4 base

stations and 10,774 occurred between supervisors and base stations. The remaining 819

detections were generated between BT-equipped supervisors, and therefore were not included in

our analyses. As stated, we needed to determine an appropriate maximum RSSI that would

suggest that supervisors were still in the presence of a given base station, and we used the 55,915

detections among the base stations to determine this maximum RSSI. Specifically, we used the

average RSSI at which base stations detected one another to identify co-location events, with the

logic that if a focal BT base station in one room can detect the BT base station in another room at a particular RSSI level, then the focal base station should also detect BT-equipped supervisors in this other room at that RSSI level. On average, the four base stations detected one another at an RSSI value of -60.08. Thus, we removed all detections from the 10,774 detections between supervisors and base stations that were less than -60 (i.e., greater than -60 in absolute value). This resulted in a sample of 5,934 BT detections between supervisors and base stations.

At this point, we calculated the number of seconds between consecutive detections of each supervisor in each location. That is, we used the timestamps of each BT detection to determine how much time had passed since a given base station (e.g., Room A's base station) had detected the BT sensor of each supervisor, as well as how much time had passed since a given supervisor sensor had detected each base station. Since the BT sensors used in this study swept for other BT sensors in 10 second intervals, we then aggregated the time each supervisor's wearable BT sensor appeared to be continuously collocated with a base station BT sensor in the rooms assigned to each team during each session. After aggregation our sample encompassed 168 observations, or 2 observations per supervisor-base station dyad – one for the supervisor's BT sensor detections of the base station, and one for the base station's BT sensor detections of the supervisor. We averaged these 2 observations to assess the time supervisors were with each team in each room (i.e., $n = 84$).

Finally, we correlated the total number of minutes derived from our calculations with the number of minutes reported by team supervisors and video coders. Table 11 provides descriptive statistics of and correlations between (a) supervisors' self-reports, (b) minutes inferred form BT-generated data, and (c) coded video data (averaged between the two coders). The supervisors' recall was quite accurate under the conditions of this study, and correlated $r = .64$ ($p < .01$) with

the coded video data. BT sensor data demonstrated a very similar and strong correlation with the

coded video data, $r = .63$ ($p < .01$). Most importantly, when included in a multiple regression

predicting coded video data, both self-reports ($\beta = .47$, $p < .01$) and BT sensor data ($\beta = .46$, $p <$

$.01$) uniquely predicted variance, with an R^2 of .58. When BT sensor data was added in a second

step, rather than in a single step with self-reports (i.e., hierarchical regression), R^2 increased by

$.18$ ($p < .05$). Thus, we conclude that BT-generated data not only detects known co-location with

a high level of accuracy but may also be used as a valuable supplement to self-reports given its

ability to predict incremental variance in time spent interacting with teams in specific locations.

Discussion

Over the last few years, wearable sensors have been subject to rapid technological

progress and wide-spread diffusion in the form of consumer electronics (e.g., fitness bands,

mobile phones). Every day, WSTs become more diverse (e.g., smart contact lenses), less

intrusive (i.e., advances in miniaturization and energy efficiency), and more affordable. As WST

communication interfaces become standardized, practitioners and organizational researchers will

have the opportunity to build affordable, custom, wearable measurement systems to collect

longitudinal, high-resolution data on human interaction in both laboratory and field settings. The

overall objective of this research was to provide construct validation evidence for the use of

wearable BT sensors in assessing relational variables. To do so we examined the construct

validity and utility of wearable BT sensor data in two studies.

The results from our first study demonstrated that BT-generated proximity data correlates

with traditional self-reported survey data, specifically self-reported friendship and task-based

relationships. As noted earlier, Eagle et al. (2009) observed correlations between self-reported

proximity and BT data that ranged from $r = .16$ to $r = .41$, which are comparable to cross-method

correlations observed for psychological constructs (e.g., Meyer et al., 2001) and our own findings. Overall, it seems that the evidence for convergent validity in the present study is favorable, and that thresholding does impact the level of convergence.

This study also provided some evidence that BT data are sensitive to the types of interactions that take place. We find that treating BT radio signal strength data as a continuous variable may add value as it could allow researchers to differentiate between different relationships (e.g., friendship, advice-giving, and advice-receiving). Our results suggest that the BT RSSI profile for interactions associated with an affective relationship (i.e., friendship) may be different from that produced by purely task-related exchanges (i.e., advice) as the interactions of the former seem to occur at closer distances. The key contribution here is that conventional approaches to BT RSSIs, which involve a threshold-based logic that (a) treats all detections above or below a particular RSSI threshold as the same and (b) provides a simple count of detections, fails to realize the full psychometric potential of BT data. Therefore, organizational researchers should consider analyzing full proximity profiles rather than binary detection counts.

Part three of this first study assessed the ability of BT data to predict individual job performance. The results from this portion of the study suggest that brokerage, a commonly-studied network variable, can be measured via BT-generated data and used to predict future performance in contexts where brokerage is critical. In the context of our study, a knowledge-intensive multiteam system working on the implementation of a large scientific installation, brokerage *is* critical and, therefore, its relationship with performance should be evident if the sensors are valid indicators of brokerage. Thus, part three of our study provides evidence that BT-generated data may allow researchers to calculate network variables that contribute to their ability to predict important workplace outcomes.

The fourth part of our first study concentrated on the potential of BT data to decompose variance in individual-level network variables and capture systematic changes in interaction patterns. The period over which we collected our data was characterized by a major, discontinuous event that provided us with a natural, quasi-experimental research setting. This discontinuous event emerged in our exploratory factor analysis and illustrated how high-resolution temporal data generated by wearable BT sensors could enhance our understanding of organizational dynamics. Thus, part four of our first study provides two important contributions. First, we discovered that a majority of variance in two ego network variables lies within-person (or unit) rather than between-persons (or units). Traditional methods for assessing networks make it a practical impossibility to explore such intra-individual (unit) variability in these constructs, and thus WSTs have the potential to open entirely new lines of research. For example, most network data are collected at one point in time and, consequently, extant research is dominated by *static* logics of structural constraint. However, continuously collected network data can move research towards a *dynamic* perspective and foster the generation of new insights into individual-level agency, dyad-level stability, and network-level periodicity (i.e., regularity and duration of representative activity states). Second, and relatedly, WSTs allow researchers to examine systematic changes in interaction patterns, as evidenced by the results from our factor analysis. Future research may use WSTs to examine changes in network activity over longer time periods to produce even finer-grained interaction patterns and interaction modes.

Finally, our second study, conducted in a heavily monitored field setting, demonstrated that BT-generated data correlated highly with videotaped, coded interactions. It should be noted that we tried to facilitate a high correlation between self-reports and this coded video data (i.e., ground truth), and determined that BT data correlated just as strongly with the coded video data

as these facilitated self-reports. This study demonstrates that BT-generated data can be used to predict unique variance in known locations, thus demonstrating how BT sensors could act as a complement to existing methods commonly used in organizational science.

Recommendations and Best Practices

Several recommendations may be derived from the findings of our work, as well as our personal experiences working with this technology. Like Chaffin et al. (2017), many of our recommendations center on planning and conducting a WST-based study. However, several additional recommendations, beyond those offered by Chaffin et al. (2017), can be gleaned from our work, particularly as it pertains to data analysis, exploration, and interpretation.

Prior to WST deployment. There are several steps we recommend researchers and practitioners engage in before commencing data collection. First, we recommend that individuals pretest the actual sensors they intend to use in a controlled environment prior to deploying them in the field. In our second study, some of our sensors suffered from software malfunctions, which resulted in a loss of data. Although we retained approximately 80% of our data, which still corresponds to a relatively favorable "response rate," a benefit of WSTs is that they can facilitate the collection of longitudinal interaction data and the loss of any such data is unfortunate.

Second, we advise researchers to focus on the measurement technology (BT) itself, rather than strictly the measurement tool (i.e., brand), when choosing a WST. Herein, we focused on BT per Chaffin et al.'s (2017) suggestions, and we examined convergence between two different BT sensors as we used Humanyze sensors in our first study and Limefy sensors in our second study. Our results suggest that BT is a useful technology and may represent a valid data collection instrument across manufacturers. However, when choosing a BT-based wearable

measurement tool, researchers need to carefully evaluate how different system and component architectures might affect the fidelity of the BT signal.

Third, we recommend that researchers and practitioners set up and stick to a reliable, consistent deployment schedule. As noted, BT records consist of sender and receiver identifiers (i.e., identification numbers). Thus, assigning and consistently attaching specific wearable sensors to specific individuals eases the examination of interaction patterns across several days while also minimizing human errors associated with the matching of sender and receiver identifiers with participant identification numbers.

Fourth, researchers and practitioners should attempt to distinguish and understand the various relationships that exist among participants and employees. The results of our discriminant validity analyses (i.e., the correlations between discrete BT RSSI values and self-reported variables, Table 2) as well as our QAP correlations suggest that (a) different types of relationships may be reflected in more frequent interactions at certain RSSI levels, (b) the prediction of various types of relationships may be differentially affected by the use of thresholding, and (c) different relationships and networks, though potentially correlated, are sufficiently distinct to warrant individual attention (given our QAP correlations between friendship and advice-giving networks ($r = .43$) and friendship and advice-receiving networks ($r = .37$)).

Fifth, although we recommend the use of self-report scales consistent with the conceptualization of the variables they are meant to capture, the use of dichotomous scales may attenuate the correlations one obtains when examining the convergence between self-report measures and BT-generated detections due to range restriction. As previously noted, the correlation between BT detections and self-reported friendship may have been attenuated given

the dichotomous nature of the friendship scale. For example, in our analyses examining discrete RSSI values the peak correlation between friendship and BT detections was $r = .35$ at an RSSI of -76. However, the use of a biserial correlation (not a *point* biserial correlation), which assumes a continuous distribution underlying the dichotomous variable (Cohen et al., 2003), increased this value to $r_b = .54$. To avoid confusion regarding convergence between self-reports and BT detections in future validation efforts (which we encourage), it is best to be consistent in the format one uses to collect self-report measures or, when it is necessary to use varying scale formats, correct and account for potential range restriction.

Finally, and related to our last recommendation, it is possible that the lower correlations we obtained between friendship and BT detections, relative to the correlations obtained between advice receiving/giving and BT detections, is partially attributable to the more psychological nature of the friendship construct. Indeed, some constructs are more behavioral and observable than others, meaning that they can be more easily perceived and detected (Carter, Carter, & DeChurch, 2018). Friendship is arguably more psychological and less observable than advice giving and receiving as the acts of giving and receiving advice represent overt behaviors that can be readily discerned and, hypothetically, more easily detected by WSTs. That is, WSTs such as BT may be more appropriate for detecting relatively behavioral phenomena as they are capable of capturing observable interaction patterns. We recommend that future researchers not only test this proposition but also consider this as a possibility in the selection of the constructs they ultimately choose to study.

During WST deployment. There are also several strategies that researchers and practitioners can use during WST deployment to ensure the success of their efforts. First, echoing suggestions made by Chaffin et al. (2017), we recommend that individuals deploy WSTs

over a relatively long window of observation. As noted, BT sensors generate detections that adhere to a Gaussian distribution (Ramadurai & Sichitiu, 2003; Sichitiu & Ramadurai, 2004), and therefore collecting observations over an extended period allows the random error associated with single BT detections to cancel out over time, thus facilitating the emergence of a "true" RSSI value.

Second, and particularly relevant to field settings, we recommend that researchers remain vigilant and catalog known changes in their research environment. Our context was characterized by a known, major transition in network activity states, namely groundbreaking (i.e., Day 0). Knowledge of this critical transition facilitated our efforts in interpreting the visible changes in our network graphs (Figures 3a through 3j). As a result, we were able to illustrate how BT-generated data can be analyzed to reflect structural regularities in network activity states, opening conceptual and empirical research opportunities largely unavailable with survey-based data collection strategies.

Finally, we recommend that researchers and practitioners do what they can to promote wearer compliance. Compliance issues, and associated missing data, can rapidly erode data quality as well as the integrity of one's results. Indeed, many important network statistics are disproportionately sensitive to missing data due the interdependence of relational data. Furthermore, the high-powered empirics associated with the impressive number of detections generated by WSTs are likely to amplify the potential impact of missing data, leading to spurious findings that inaccurately reflect relationships and interactions among individuals. Granted, we must balance personal privacy with our desire for complete data, but the communication of well-intentioned, scientifically-oriented motivation to participants could help (as well as the guaranteed anonymization of data, post data collection).

After WST deployment. Where our recommended practices deviate most from those provided by Chaffin et al. (2017) is in the processing, exploration, and analysis of data, given our field contexts. First, in the processing of data, we recommend that researchers avoid using preset RSSI thresholds to determine which detections represent "meaningful" interaction opportunities and which detections do not (i.e., which detections represent distances too great to constitute an interaction opportunity). The results addressing our first research question (Table 1), that is, BT's convergence with self-reported relational measures, suggest that the highest convergence between BT-generated data and self-reports of relational measures occur at RSSI thresholds greater than the recommended -80 (Olguin et al., 2009). Relatedly, the results of all our ERGM analyses indicate that thresholding does not enhance our ability to predict either advice-oriented or friendship relationships, regardless of whether we included information pertaining to the formal organizational structure (i.e., functional background), given increases in AICs and BICs. Perhaps there are research contexts in which thresholding makes practical and/or theoretical sense (e.g., in small laboratory suites where the outcome of interest is time spent in specific locations), but blindly utilizing an RSSI threshold based upon precedent or common practice is problematic as it may limit one's ability to predict outcomes of interest.

Second, and similarly related to the processing of data, we recommend that researchers and practitioners screen for outliers within their registered BT detections. That is, researchers and practitioners should ensure that their WST-generated data are realistic by examining the detection records themselves. For example, and as previously noted, we discovered and purged detections that represented improbable interactions based upon the timestamp of the detection record. In total, we removed 3,762 detections (1.7% of all data generated) as these detections were generated after-hours, when the sensors were in the possession of the research team. This

was likely due to participants failing to turn the sensors off or inadvertent activation after they were collected. Inclusion of these detections could have led us to misguided conclusions, and thus it was necessary to purge them. Although this may seem just common sense, our point is to remember that humans (both research participants and research assistants) are error-prone, and this is something that could easily be lost when getting caught up in processing the "big data" that is collected by WSTs.

Third, we recommend that researchers and practitioners take a somewhat exploratory approach to their data analysis when it comes to WST data. As noted, we are still at the validation stage with this technology and, thus, there is still much to learn before we engage in significant substantive research. Open-mindedness with, and exploration of, WST-generated data may provide insights that might be missed by a researcher heavily focused on an initial plan. For example, we discovered that detections at stronger (i.e., less negative) RSSI values are more highly correlated with self-reports of friendship than self-reports of advice-giving and receiving (Table 2). This means that more frequent BT detections registered at RSSI values indicating closer proximity (e.g., -76) may serve as an indicator of a friendship relationship between two individuals whereas frequent BT detections registered at RSSI values indicating further proximity (e.g., -83) may serve as a better indicator of an advice-based relationship between two colleagues.

Relatedly, we recommend examining WST data at various levels of analysis, including the dyadic, individual, and whole network levels of analysis, as we did in this investigation. Organizations are inherently multilevel entities (Kozlowski & Klein, 2000) and, thus, the exploration of WST-generated data at each level represents a unique opportunity to evaluate its potential to contribute to future research. Similarly, though perhaps more related to the planning

of the project, we do not yet know what observation window is appropriate for capturing a particular organizational phenomenon or event. In our first study, we used an observation window of several days spanning a major organizational event (i.e., groundbreaking). However, other researchers may find that a shorter (i.e., several hours) or longer (i.e., several months) observation window is sufficient and/or necessary given their phenomenon or level of interest; the appropriate observation window may vary with the level at which the phenomenon of interest resides (i.e., dyadic, individual, or whole network level). Thus, we recommend that researchers identify and use a theoretical logic to drive their approach.

Fourth, we recommend that researchers and practitioners use this data to generate widely-studied network measures such as brokerage. These measures are not only more objective than those derived from surveys but may also predict outcomes of practical interest (i.e., performance). Indeed, our BT-generated brokerage variable predicted individual performance above and beyond self-reported measures of advice-giving, advice-receiving, and friendship, despite a relatively small sample size ($n = 31$ at the individual level of analysis). Future researchers should use BT-generated network variables to predict outcomes of interest, examine potential antecedents and consequences of these objective network measures, and assess their convergence with and divergence from self-reported measures of network variables (i.e., are the antecedents of network variables identified in survey research also predictive of network variables derived from BT detections?).

Fifth, and finally, researchers and practitioners should examine *variance* in BT-generated network measures, ensuring that the observation window over which they collect their data (i.e., hours, days, months) corresponds, theoretically, to their phenomenon of interest (you miss the variance you do not capture). The results of our variance decomposition of network measures at

the individual level of analysis suggested that a majority of variance resides within individuals rather than between individuals. When it comes to conceptualizing one's position in the network, researchers have been unable to explore the extent to which any one person's position in a network is contingent upon the work configuration associated with different days due to the limitations of self-reported network data, and we find that a great deal of variance is lost by aggregating to a composite level (i.e., across all nine days of data collection in Study 1). We have emphasized before, and feel the need to emphasize again, that no prior organizational study has been able to examine "daily networks" like this, and, hence, extant research may be missing where the majority of variance lies in network structures. Thus, the use of WSTs may not only open entirely new lines of research but may revolutionize our understanding of networks altogether.

Limitations

The findings and implications of this study should be interpreted with its limitations in mind. The most notable limitations associated with this research are (a) the unique nature of the research contexts and (b) the relatively small number of participants. However, these samples were chosen for two reasons. First, the work context and specialization of the participants in the first study facilitated a high level of interaction and task interdependence, which ensured a large number of BT detections. In terms of the second study, the small number of participants allowed us to monitor and accurately code videotaped interactions between supervisors and teams, which allowed us to maximize correlations between self-reports, BT data, and coded video data. Second, and relatedly, the use of BT sensors in organizational research is nascent and, thus, we were required to restrict the number of participants to a manageable size. Future developments of WSTs may alleviate some of the logistical demands involved in their deployment.

Another potential concern with this study is that the instability of RSSI can make it difficult to establish a definite relationship between distance and RSSI (Heurtefeux & Valois, 2012). However, it is important to reemphasize that we collected data for relatively long periods of time in both studies (i.e., several weeks), in settings where tasks were highly interdependent across functional backgrounds (Study 1) and physical locations (Study 2). Accordingly, the large number of BT detections we obtained should enable us to leverage this imperfect measure to approximate actual proximity. In support of this, prior researchers demonstrated that RSSI can be used as a distance measurement in various indoor and outdoor locations (Eagle et al., 2009; Kumar, Reddy, & Varma, 2009; Liu, Jiang, & Striegel, 2014; Naya, Noma, Ohmura, & Kogure, 2005; Zhang, Zhang, Ying, & Gao, 2009).

Additionally, we cannot say with absolute certainty that the validity evidence we present here will generalize to all other relationship and network types, or to other constructs typically examined with a network lens. What we provided was initial evidence that, over extended observation periods in field settings, BT-sensor generated proximity data could potentially help researchers and practitioners identify and discriminate between friendship and task-based relationships (i.e., advice relationships), as well as detect (a) major transitions in network activity states, (b) time spent by individuals in various locations, and (c) interactions among organizational actors, more generally. We did not provide evidence that BT sensors can be used to identify relationships in alternative, specific networks (e.g., developmental and knowledge networks; Dobrow, Chandler, Murphy, & Kram, 2012; Phelps et al., 2012), nor did we provide evidence that BT sensors may be used to capture all constructs that may be examined with a network lens, such as social capital (Li, 2013) and trust (Ferrin, Dirks, & Shah, 2006).

We cannot assume that the validation evidence that we provide herein will generalize to all alternative networks and constructs (and their associated interaction patterns), and we do not yet know what constitutes an appropriate BT data collection period for the discrimination among various relationship types. Therefore, much more validation work needs to be done by psychologists before we can definitively say that BT sensor data reflects specific psychological constructs not examined here (e.g., trust, distrust) as well as those examined here (i.e., friendship, advice-oriented relationships). That is, future research should not only attempt to replicate the validation evidence we provide (i.e., BT data's validity with respect to friendship, advice-oriented relationships, and interactions, more generally) but also attempt to validate BT sensors in contexts associated with alternative types of networks and relationships.

Finally, and given the evidence on individual and whole network variability presented herein, one may be concerned that there was a three-month gap between the collection of self-reported friendship, advice-giving, and advice-receiving and the collection of BT detections in Study 1. Specifically, one may be concerned that the networks captured by the BT data and the networks captured by self-reports represent notably different networks given our evidence that networks vary and evolve over time. Although there may be some degradation in convergent, discriminant, and predictive validity attributable to the temporal delay, the level of this degradation did not prevent us from obtaining statistically significant results. Nevertheless, future research should attempt to capture self-reports and BT data closer together in time, not only to see whether stronger effects can be obtained (as measurements closer together in time should, presumably, result in stronger correlations than those we obtained) but also to document the potential degradation associated with increasingly larger time lags (e.g., is the correlation between self-reported relationships and BT-generated data greater as the time lag decreases? By

how much does the correlation decrease as the lag increases?). Though this may be somewhat difficult considering the labor intensive nature of the repeated collection of network surveys and the inaccuracies associated with participant recall of instances of interaction during specific timeframes (Freeman, 1992; Marsden, 1990), findings that speak to the degradation of convergent validity resulting from increasingly large time lags would be both informative and supportive of our arguments regarding network variability and, relatedly, evolution.

Conclusion

When one considers how social and behavioral scientists seem to readily adopt new technologies (i.e., email, social media platforms, panoramic cameras), it seems inevitable that the use of WSTs will someday become routine practice. Although we are still far from that day, as market forces and scientific progress continue to facilitate the development of these technologies social and behavioral scientists should look for opportunities to exploit the increased connectedness and granularity that WSTs have to offer. In an effort to speed the arrival of that day, the present investigation represents the first attempt to perform a traditional psychometric construct validation study that assesses the utility of WST-generated data (i.e., Bluetooth), with an emphasis on how it can be used to leverage and contribute to the extant knowledge base. This study represents an initial step down a long road of research that may ultimately revolutionize data collection, analytical procedures, and research objectives of social and behavioral scientists.

References

Alshamsi, A., Pianesi, F., Lepri, B., Pentland, A., & Rahwan, I. (2015). Beyond contagion:

Reality mining reveals complex patterns of social influence. *PLoS One, 10*. doi:

10.1371/journal.pone.0135740.

Bernard, H. R., Killworth, P. D., & Sailer, L. (1981). Summary of research on informant

accuracy in network data and the reverse small world problem. *Connections, 4*, 11-25.

Blau, P. M., Blum, T. C., & Schwartz, J. E. (1982). Heterogeneity and intermarriage. *American

Sociological Review, 47*, 45-62.

Bliese, P. D. (2000). Within-group agreement, non-independence, and reliability: Implications

for data aggregation and analysis. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel

theory, research, and methods in organizations* (pp. 349-381), San Francisco, CA:

Jossey-Bass.

Borgatti, S. P., Everett, M. G. & Freeman, L. C. (2002). *UCInet 6 for Windows: Software for

social network analysis*. Harvard, MA: Analytic Technologies.

Burt, R. S. (1992). *Structural holes: The social structure of competition*. Cambridge, MA:

Harvard University Press.

Burt, R. S. (2004). Structural holes and good ideas. *American Journal of Sociology, 110*, 349-

399.

Carter, N. T., Carter, D. R., & DeChurch, L. A. (2018). The implications of observability for the

theory and measurement of emergent team phenomena. *Journal of Management, 44*,

1398-1425.

Cattuto, C., Van den Broeck, W., Barrat, A., Colizza, V., Pinton, J-F., Vespignani, A. (2010). Dynamics of person-to-person interactions from distributed RFID sensor networks. *PLoS One, 5*. doi: 10.1371/journal.pone.0011596

Chaffin, D., Heidl, R., Hollenbeck, J. R., Howe, M., Yu, A., Voorhees, C., & Calantone, R. (2017). The promise and perils of wearable sensors in organizational research. *Organizational Research Methods, 20*, 3-31.

Clement, J., Shipilov, A., & Galunic, D. C. (in press). Brokerage as a public good: The externalities of network hubs for different formal roles in creative organizations. *Administrative Science Quarterly*.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. New York, NY: Routledge.

Collopy, F. (1996). Biases in retrospective self-reports of time use: An empirical study of computer users. *Management Science, 42*, 758-767.

Crawford, E. R., & LePine, J. A. (2013). A configural theory of team processes: Accounting for the structure of taskwork and teamwork. *Academy of Management Review, 38*, 32-48.

Dalal, R., Lam, H., Weiss, H., Welch, E., & Hulin, C. (2009). A within-person approach to work behavior and performance: Concurrent and lagged citizenship-counterproductivity associations, and dynamic relationships with affect and overall job performance. *Academy of Management Journal*, *52*, 1051–1066.

Dobrow, S. R., Chandler, D. E., Murphy, W. M., & Kram, K. E. (2012). A review of developmental networks: Incorporating a mutuality perspective. *Journal of Management, 38*, 210-242.

Dulebohn, J. H., Bommer, W. H., Liden, R. C., Brouer, R. L., & Ferris, G. R. (2012). A meta-
analysis of antecedents and consequences of leader-member exchange: Integrating the
past with an eye toward the future. *Journal of Management, 38*, 1715-1759.

Eagle, N., Pentland, A. S., & Lazer, D. (2009). Inferring friendship network structure by using
mobile phone data. *Proceedings of the National Academy of Sciences of the United States
of America, 106*, 15274-15278.

Ebadi, Y. M., Utterback, J. M. (1984). The effects of communication on technological
innovation. *Management Science, 30*, 572-585.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use
of exploratory factor analysis in psychological research. *Psychological Methods, 4*, 272-
299.

Ferrin, D. L., Dirks, K. T., & Shah, P. P. (2006). Direct and indirect effects of third-party
relationships on interpersonal trust. *Journal of Applied Psychology, 91,* 870-883.

Ferro, E., & Potorti, F. (2005). Bluetooth and Wi-Fi wireless protocols: A survey and a
comparison. *IEEE Wireless Communications, 12*, 12-26.

Finnerty, A. N., Kalimeri, K., & Pianesi, F. (2014). Towards happier organisations:
Understanding the relationship between communication and productivity. In L. M. Aiello
& D. McFarland (Eds.), *Social informatics* (pp. 462-477). Cham, Switzerland: Springer
International Publishing.

Fleming, L., Mingo, S., & Chen, D. (2007). Collaborative brokerage, generative creativity, and
creative success. *Administrative Science Quarterly, 52,* 443-475.

Fournet, J., & Barrat, B. (2014). Contact patterns among high school students. *PLoS One, 9*. doi:
10.1371/journal.pone.0107878

Freeman, L. C. (1992). Filling in the blanks: A theory of cognitive categories and the structure of social affiliation. *Social Psychology Quarterly*, *55*, 118-127.

Gibbons, D. E. (2004). Friendship and advice networks in the context of changing professional values. *Administrative Science Quarterly, 49*, 238-262.

Gosling, S. D., John, O. P., Craik, K. H., & Robins, R. W. (1998). Do people know how they behave? Self-reported act frequencies compared with on-line codings by observers. *Journal of Personality and Social Psychology, 74*, 1337-1349.

Haartsen, J., Naghshineh, M., Inouye, J., Joeressen, O. J., & Allen, W. (1998). Bluetooth: Vision, goals, and architecture. *ACM SIGMOBILE Mobile Computing and Communications Review, 2,* 38-45.

Handcock, M. S., Hunter, D. R., Butts, C. T., Goodreau, S. M., & Morris, M. (2008). statnet: Software tools for the representation, visualization, analysis and simulation of network data. *Journal of Statistical Software, 24,* 1-11.

Heurtefeux, K., & Valois, F. (2012). Is RSSI a good choice for localization in wireless sensor network? *2012 IEEE 26th International Conference on Advanced Information Networking and Applications,* doi: 10.1109/AINA.2012.19

Ibarra, H., & Andrews, S. B. (1993). Power, social influence, and sense making: Effects of network centrality and proximity on employee perceptions. *Administrative Science Quarterly, 38*, 277-303.

James, L. R. (1982). Aggregation bias in estimates of perceptional agreement. *Journal of Applied Psychology, 67,* 219-229.

Johnson, R. E., Lanaj, K., & Barnes, C. M. (2014). The good and bad of being fair: Effects of procedural and interpersonal justice behaviors on regulatory resources. *The Journal of*

Applied Psychology, *99*, 635–650.

Kalimeri, K. (2013). Towards a dynamic view of personality: Multimodal classification of

personality states in everyday situations. *Proceedings of the 15th ACM on International*

conference on multimodal interaction, doi: 10.1145/2522848.2532195

Kamada, T., & Kawai, S. (1989). An algorithm for drawing general undirected graphs.

Information Processing Letter, *31*, 7-15

Kammar, D., McNutt, G., Senese, B., & Bray, J. (2002). *Bluetooth application developer's*

guide: The short range interconnect solution. Rockland, MA: Syngress Publishing, Inc.

Kozlowski, S. W. J, & Klein, K. J. (2000). A multilevel approach to theory and research in

organizations: Contextual, temporal, and emergent processes. In K. J. Klein & S. W. J.

Kozlowski (Eds.), *Multilevel theory, research and methods in organizations:*

Foundations, extensions, and new directions (pp. 3-90), San Francisco, CA: Jossey-Bass

Kumar, P., Reddy, L., & Varma, S. (2009). Distance measurement and error estimation scheme

for RSSI based localization in wireless sensor networks. *2009 Fifth International*

Conference on Wireless Communication and Sensor Networks (WCSN), doi:

10.1109/WCSN.2009.5434802

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A-L., Brewer, D., Christakis, N.,

Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., & Van

Alstyne, M. (2009). Computational social science. *Science, 323*, 721-723.

Li, M. (2013). Social network and social capital in leadership and management research: A

review of causal methods. *Leadership Quarterly, 24*, 638-665.

Liu, S., Jiang, Y., & Striegel, A. (2014). Face-to-face proximity estimation using Bluetooth on

smartphones. *IEEE Transactions on Mobile Computing, 13*, 811-823.

Marsden, P. V. (1990). Network data and measurement. *Annual Review of Sociology, 16*, 435-
    465.

Mastrandrea, R., Fournet, J., & Barrat, A. (2015). Contact patterns in a high school: A
    comparison between data collected using wearable sensors, contact diaries and friendship
    surveys. *PLoS One, 10,* doi: 10.1371/journal.pone.0136497

Mathieu, J. E., Hollenbeck, J. R., van Knippenberg, D., & Ilgen, D. R. (2017). A century of work
    groups in the Journal of Applied Psychology. *Journal of Applied Psychology*, *102*, 452–467.

Matta, F. K., Scott, B. A., Colquitt, J. A., & Passantino, L. G. (2017). Is consistently unfair better
    than sporadically fair? An investigation of justice variability and stress. *Academy of
    Management Journal, 60*, 743–770.

McFadyen, M. A., Semadeni, M., Cannella, A. A. (2009). Value of strong ties to disconnected
    others: Examining knowledge creation in biomedicine. *Organization Science, 20*, 552-564.

Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., Eisman, E. J.,
    Kubiszyn, T., & Reed, G. M. (2001). Psychological testing and psychological
    assessment: A review of evidence and issues. *American Psychologist*, *56,* 128-165.

Morrison, E. W. (2002). Newcomer's relationships: The role of social networks ties during
    socialization. *Academy of Management Journal, 45,* 1149–1160.

Morrison, E. W. (2010). From the editors: OB in AMJ: What is hot and what is not? *Academy of
    Management Journal, 53*, 932-936.

Muthén, L. K., & Muthén, B. O. (1998-2011). *Mplus user's guide*. Los Angeles, CA: Muthén &
    Muthén.

Naya, F., Noma, H., Ohmura, R., & Kogure, K. (2005). Bluetooth-based indoor proximity

     sensing for nursing context awareness. *Ninth IEEE International Symposium on*

     Wearable Computers (ISWC '05), doi: 10.1109/ISWC.2005.13

Nunnally, J. (1978). *Psychometric methods*. New York, NY: McGraw-Hill.

Olguin, D., Waber, B., Kim, T., Mohan, A., Ara, K., & Pentland, A. (2009). Sensible

     organizations: Technology and methodology for automatically measuring organizational

     behavior. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 39*,

     43–55.

Onnela, J-K., Waber, B. N., Pentland, A., Schnorf, S., & Lazer, D. (2014). Using sociometers to

     quantify social interaction patterns. *Scientific Reports, 4*, 1-9.

Panisson, A., Barrat, A., Cattuto, C., Van den Broeck, W., Ruffo, G., & Schifanello, R. (2012).

     On the dynamics of human proximity for data diffusion in ad-hoc networks. *Ad Hoc*

     Networks, 10, 1532-1543.

Pentland, A. (2012). The new science of building great teams. *Harvard Business Review, 90,* 3-

     11.

Phelps, C., Heidl, R., & Wadhwa, A. (2012). Knowledge, networks, and knowledge networks: A

     review and research agenda. *Journal of Management, 38*, 1115-1166.

Ramadurai, V., & Sichitiu, M. L. (2003). Localization in wireless sensor networks: A

     probabilistic approach. *Published 2003 in International Conference on Wireless*

     Networks, 275-281.

Reagans, R., Zuckerman, E., & McEvily, B. (2004). How to make the team: Social networks vs.

     demography as criteria for designing effective teams. *Administrative Science Quarterly,*

     49, 101-133.

Rosenthal, R., & Rosnow, R. (2008). *Essentials of behavioral research: Methods and data analysis.* New York, NY: McGraw-Hill.

Schwab, D. (1980). Construct validity in organizational behavior. In L. L. Cummings & B. M. Staw (Eds.), *Research in organizational behavior (Vol. 2)* (pp. 2-43). Greenwhich, CT: JAI Press.

Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist, 54*, 93-105.

Scott, B. A., Garza, A. S., Conlon, D. E., & Kim, Y. J. (2014). Why do managers act fairly in the first place? A daily investigation of "hot" and "cold" motives and discretion. *Academy of Management Journal*, *57*, 1571–1591.

Seibert, S. E., Kraimer, M. L., & Liden, R. C. (2001). A social capital theory of career success. *Academy of Management Journal, 44*, 219-237.

Sekara, V., & Lehmann, S. (2014). The strength of friendship ties in proximity sensor data. *PLOS One, 9,* doi: 10.1371/journal.pone.0100915

Sekara, V., Stopczynski, A., & Lehmann, S. (2016). Fundamental structures of dynamic social networks. *Proceedings of the National Academy of Sciences of the United States of America, 113*, 9977-9982.

Sias, P. M., & Cahill, D. J. (1998). From coworkers to friends: The development of peer friendships in the workplace. *Western Journal of Communication, 62*, 273-299.

Sichitiu, M. L., & Ramadurai, V. (2004). Localization of wireless sensor networks with a mobile beacon. *2004 IEEE International Conference on Mobile Ad-hoc and Sensor Systems,* 174-183.

Snijders, T. A., Pattison, P. E., Robins, G. L., & Handcock, M. S. (2006). New specifications for

    exponential random graph models. *Sociological Methodology, 36,* 99-153.

StataCorp. (2013). *Stata Statistical Software: Release 13*. College Station, TX: StataCorp LP.

Stewart, G. L., & Nandkeolyar, A. K. (2007). Exploring how constraints created by other people

    influence intraindividual variation in objective performance measures. *Journal of Applied*

    Psychology, *92*, 1149–1158.

Stöber, J., Dette, D. E., & Musch, J. (2002). Comparing continuous and dichotomous scoring of

    the balanced inventory of desirable resonding. *Journal of Personality Assessment, 78*, 370-

    389.

Stopczynski, A., Sekara, V., Sapiezynski, P., Cuttone, A., Madsen, M. M., Larsen, J. E., &

    Lehmann, S. (2014). Measuring large-scale social networks with high resolution. *PLoS One,*

    9, doi: 10.1371/journal.pone.0095978

To, M. L., Fisher, C. D., Ashkanasy, N. M., & Rowe, P. A. (2012). Within-person relationships

    between mood and creativity. *Journal of Applied Psychology*, *97*, 599–612.

Tsui, A. S., Pearce, J. L., Porter, L. W., & Tripoli, A. M. (1997). Alternative approaches to the

    employee-organization relationship: Does investment in employees pay off? *Academy of*

    Management Journal, 40, 1089-1121.

Wang, X. H., & Iqbal, M. (2006). Bluetooth: Opening a blue sky for healthcare. *Mobile*

    Information Systems, 2, 151-167.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications.*

    Cambridge, England: Cambridge University Press.

York, R. C., Bollen, G., Compton, C., Crawford, A. C., Doleans, M., & Glasmacher, T., Hartung,

    W., Marti, F., Popielarski, J. J., Vincent, J., Wlodarczak, X., Wu, A., Zeller, Q., & Zhao,

Q. (2009). FRIB: A new accelerator facility for the production of rare isotope beams.

Proceedings of SRF 2009, 888-894.

Zhang, J. W., Zhang, L., Ying, Y., & Gao, F. (2009). Research on distance measurement based

on RSSI of ZigBee. *Chinese Journal of Sensors and Actuators, 2*, 139-142.

Table 1

Correlations between Cumulative Bluetooth Signal Detections and Self-Reported Variables

RSSI	Friendship	Advice-Giving	Advice-Receiving
-69 (1,284)	.236	.381	.403
-70 (2,624)	.254	.385	.411
-71 (3,836)	.261	.392	.419
-72 (5,375)	.271	.409	.431
-73 (9,706)	.285	.428	.449
-74 (10,767)	.289	.426	.450
-75 (13,685)	.296	.430	.455
-76 (16,837)	.309	.438	.463
-77 (21,202)	.316	.445	.473
-78 (28,707)	.325	.452	.484
-79 (34,595)	.330	.454	.491
-80 (45,996)	.338	.461	.503
-81 (54,639)	.341	.466	.508
-82 (66,234)	.341	.470	.515
-83 (92,535)	**.344**	**.479**	.525
-84 (101,151)	.343	.479	.526
-85 (117,456)	.342	.479	.529
-86 (132,129)	.340	.478	.529
-87 (147,870)	.340	.478	.530
-88 (174,123)	.339	.477	.531
-89 (182,221)	.338	.476	**.532**
-90 (190,547)	.337	.474	.532
-91 (202,098)	.337	.473	.532

Note. $n = 992$ (dyads) at each threshold. Number of detections generated at each threshold in parentheses. All correlations significant at $p < .01$. Average standard errors for friendship, advice-giving, and advice-receiving are .030, .028, and .028, respectively. Peak correlations bolded and underlined.

Table 2

Correlations between Discrete Bluetooth Signal Detections and Self-Reported Variables

RSSI	Friendship	Advice-Giving	Advice-Receiving
-69 (1,284)	.236	.382	.403
-70 (1,340)	.257	.367	.397
-71 (1,212)	.253	.374	.401
-72 (1,539)	.278	.421	.429
-73 (4,331)	.289	.432	.450
-74 (1,061)	.287	.358	.401
-75 (2,918)	.303	.415	.443
-76 (3,152)	**.349**	.443	.466
-77 (4,365)	.323	.440	.479
-78 (7,505)	.332	.447	.492
-79 (5,888)	.323	.427	.488
-80 (11,401)	.340	.455	.507
-81 (8,643)	.330	.449	.497
-82 (11,595)	.316	.455	**.508**
-83 (26,301)	.323	**.461**	.507
-84 (8,616)	.298	.436	.488
-85 (16,305)	.298	.433	.497
-86 (14,673)	.292	.421	.474
-87 (15,741)	.298	.426	.480
-88 (26,253)	.296	.416	.474
-89 (8,098)	.273	.387	.459
-90 (8,326)	.253	.336	.452
-91 (11,551)	.242	.286	.379

Note. n = 992 (dyads) at each threshold. Number of detections generated at each threshold in parentheses. All correlations significant at $p < .01$. Average standard errors for friendship, advice-giving, and advice-receiving are .030, .029, and .028, respectively. Peak correlations bolded and underlined.

Table 3

Exponential Random Graph Models of Directed Friendship Network using BT Detections as Edge Covariate (Model fitted using MCMC): Test of Predictive Value of Thresholding at RSSI > -80

Fit for ERGM without Thresholding

Parameter	Estimate	SE
Density	-2.668*	0.135
Reciprocity	1.576*	0.311
Daily Average BT Detections (Unrestricted)	0.017*	0.0024

AIC: 711.5   BIC: 726.2

Fit for ERGM with Thresholding

Parameter	Estimate	SE
Density	-2.580*	0.132
Reciprocity	1.661*	0.311
Daily Average BT Detections (Restricted > -80)	0.065*	0.010

AIC: 723.4   BIC: 738

Note. Assumes no information with respect to formal organizational structure.

*p < .05

Table 4

Exponential Random Graph Models of Directed Friendship Network using BT Detections as Edge Covariate (Model fitted using MCMC): Test of Predictive Value of Thresholding at RSSI > -80 within a Known Formal Organizational Structure

Fit for ERGM without Thresholding

Parameter	Estimate	SE
Density	-3.451*	0.240
Reciprocity	1.093*	0.332
Functional Homophily	1.610*	0.267
Business Support Team	0.106	0.183
Scientific Team	0.634*	0.156
Leadership Team	-0.002	0.995
Daily Average BT Detections (Unrestricted)	0.01*	0.002

AIC: 659.2   BIC: 693.5

Fit for ERGM with Thresholding

Parameter	Estimate	SE
Density	-3.307*	0.224
Reciprocity	1.116*	0.335
Functional Homophily	1.688*	0.254
Business Support Team	-0.008	0.174
Scientific Team	0.507*	0.150
Leadership Team	-0.082	0.776
Daily Average BT Detections (Restricted > -80)	0.034*	0.002

AIC: 661.5   BIC: 695.4

Note. Given information on formal structure.

*$p < .05$

Table 5

Exponential Random Graph Models of Directed Advice-Giving Network using BT Detections as Edge Covariate (Model fitted using MCMC): Test of Predictive Value of Thresholding at RSSI > -80

Fit for ERGM without Thresholding

Parameter	Estimate	*SE*
Density	-1.586186*	0.121439
Reciprocity	2.052295*	0.222042
Daily Average BT Detections (Unrestricted)	0.025114*	0.002976

AIC: 1104   BIC: 1118

Fit for ERGM with Thresholding

Parameter	Estimate	*SE*
Density	-1.56602*	0.12096
Reciprocity	2.14437*	0.22532
Daily Average BT Detections (Restricted > -80)	0.10382*	0.01391

AIC: 1133   BIC: 1148

Note. Assumes no information with respect to formal organizational structure.

*p < .05

Table 6

Exponential Random Graph Models of Directed Advice-Giving Network using BT Detections as Edge Covariate (Model fitted using MCMC): Test of Predictive Value of Thresholding at RSSI > -80 within a Known Formal Organizational Structure

Fit for ERGM without Thresholding

Parameter	Estimate	*SE*
Density	-1.811711*	0.180137
Reciprocity	1.911255*	0.228771
Functional Homophily	0.958569*	0.186034
Business Support Team	0.091587	0.134630
Scientific Team	0.001711	0.123393
Leadership Team	0.782271*	0.179695
Daily Average BT Detections (Unrestricted)	0.014457*	0.003405

AIC: 1074   BIC: 1108

Fit for ERGM with Thresholding

Parameter	Estimate	*SE*
Density	-1.69385*	0.17943
Reciprocity	1.93031*	0.23296
Functional Homophily	1.12202*	0.17757
Business Support Team	-0.01817	0.13343
Scientific Team	-0.10621	0.12675
Leadership Team	0.71863*	0.17372
Daily Average BT Detections (Restricted > -80)	0.05020*	0.01397

AIC: 1078   BIC: 1112

Note. Given information on formal structure.

p < .05

Table 7

Exponential Random Graph Models of Directed Advice-Receiving Network using BT Detections as Edge Covariate (Model fitted using MCMC): Test of Predictive Value of Thresholding at RSSI > -80

Fit for ERGM without Thresholding

Parameter	Estimate	SE
Density	-1.674612*	0.126810
Reciprocity	2.258182*	0.237821
Daily Average BT Detections (Unrestricted)	0.034460*	0.003842

AIC: 1030   BIC: 1044

Fit for ERGM with Thresholding

Parameter	Estimate	SE
Density	-1.69824*	0.12653
Reciprocity	2.36210*	0.23131
Daily Average BT Detections (Restricted > -80)	0.15672*	0.01943

AIC: 1058   BIC: 1072

Note. Assumes no information with respect to formal organizational structure.

*p < .05

Table 8

Exponential Random Graph Models of Directed Advice-Receiving Network using BT Detections as Edge Covariate (Model fitted using MCMC): Test of Predictive Value of Thresholding at RSSI > -80 within a Known Formal Organizational Structure

Fit for ERGM without Thresholding

Parameter	Estimate	SE
Density	-1.756360*	0.188456
Reciprocity	2.160357*	0.238464
Functional Homophily	0.820047*	0.196386
Business Support Team	-0.018820	0.129496
Scientific Team	-0.058307	0.130642
Leadership Team	0.635753*	0.173710
Daily Average BT Detections (Unrestricted)	0.022210*	0.004332

AIC: 1009   BIC: 1044

Fit for ERGM with Thresholding

Parameter	Estimate	SE
Density	-1.62300*	0.18295
Reciprocity	2.16691*	0.23252
Functional Homophily	1.00659*	0.18503
Business Support Team	-0.16420	0.13031
Scientific Team	-0.18650	0.12708
Leadership Team	0.53465*	0.17245
Daily Average BT Detections (Restricted > -80)	0.09003*	0.01989

AIC: 1010   BIC: 1045

Note. Given information on formal structure.

*p < .05

Table 9

Correlations among Variables at the Individual Level of Analysis

Variables	Mean	SD	(1)	(2)	(3)	(4)
Performance	4.29	0.60				
Brokerage	0.23	0.15	0.39*			
Friendship	3.77	4.05	.20	.20		
Advice-Giving	2.75	0.68	.06	.21	.26	
Advice-Receiving	3.19	0.59	-.19.	.29	.09	.55*

Note: *n* = 31 (individual-level). Brokerage variable normalized by number of brokerage opportunities.

*p < .05

Table 10

Exploratory Factor Analytic Results

Day	Public Site Visitation (Pre-Groundbreaking)	Normal Operational Mode (Post-Groundbreaking)
-6	.749	
-4	.538	
+2		.600
+4		.761
+8		.531
+10		.414
+14		.586
+16		.494
+18		.486

Note. Days numbered to reflect the number of days before and after groundbreaking (e.g., Day -6 occurred 6 days before groundbreaking, Day +4 occurred 4 days after groundbreaking). Factor loadings less than .3 not provided for ease of presentation. Varimax rotation used.

Table 11

Correlations between Time Spent in Suite Locations implied by Team Supervisors' Self-Reports, Video Coding, and BT Sensors

Variables	Mean	SD	(1)	(2)
Video Coding	19.86	15.70		
Self-Reports	22.87	15.76	.64	
BT Sensors	07.04	10.63	.63	.36

Note: n = 84 (person-team dyads). All correlations significant at p < .01.
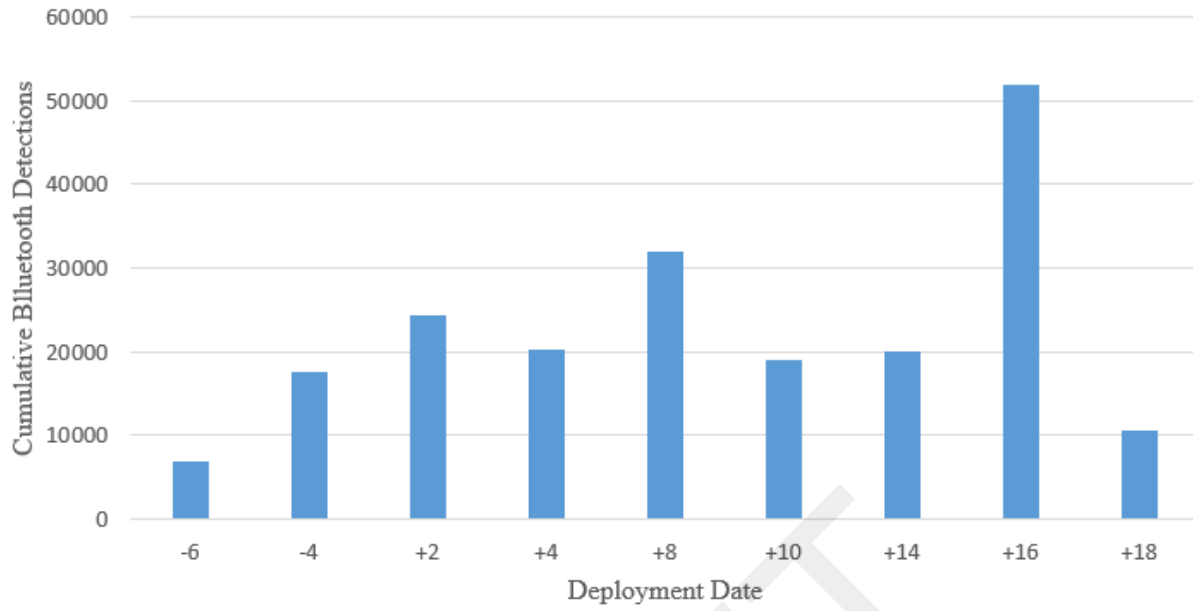
Figure 1. Cumulative daily Bluetooth detections (-69 <= RSSI <= -91) across 9 deployment days
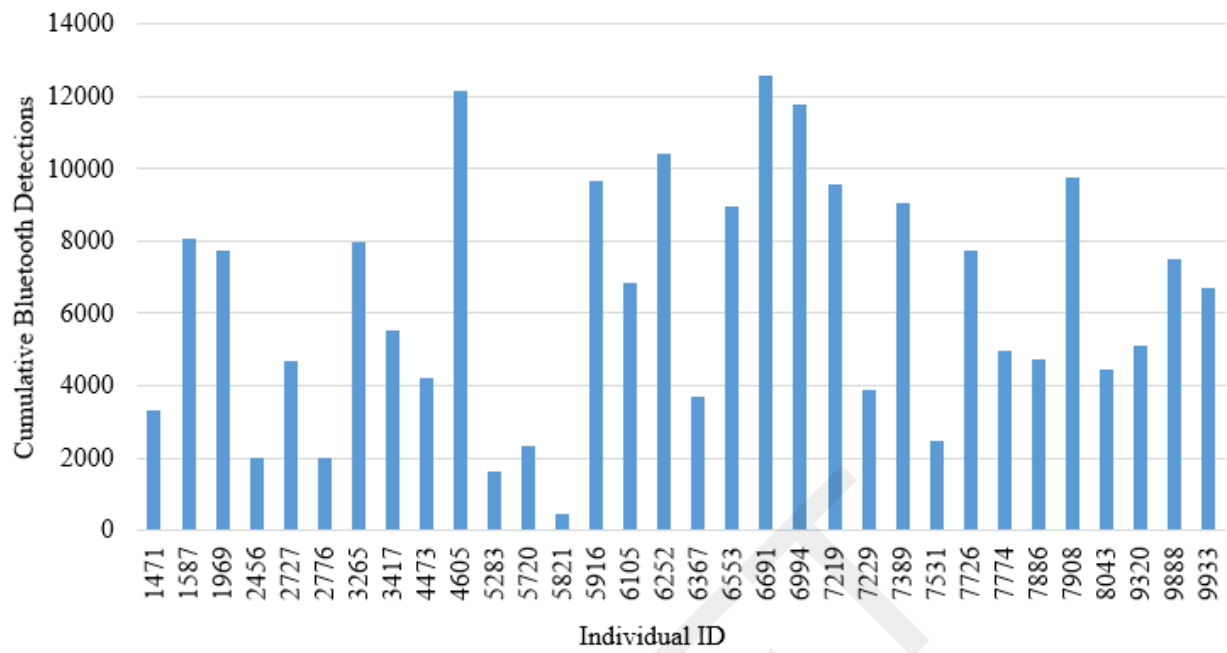
Figure 2. Cumulative participant Bluetooth detections (-69 <= RSSI <= -91) across 9 deployment days
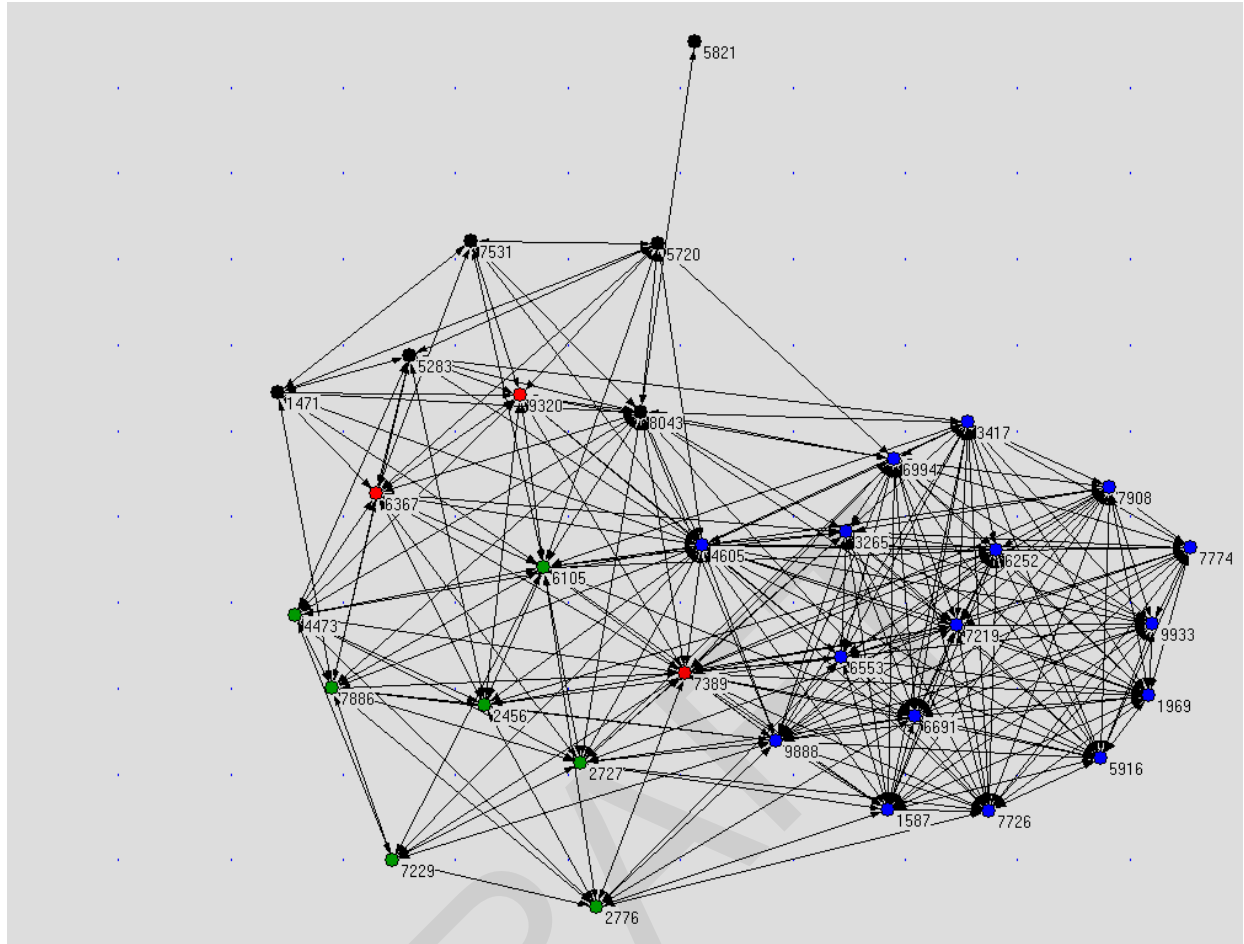
Red = FRIB Leadership; Blue = Engineering; Green = Scientists; Black = Business Support

Figure 3a. Days -6 through Day +18: Cumulative Bluetooth detections across 9 days.
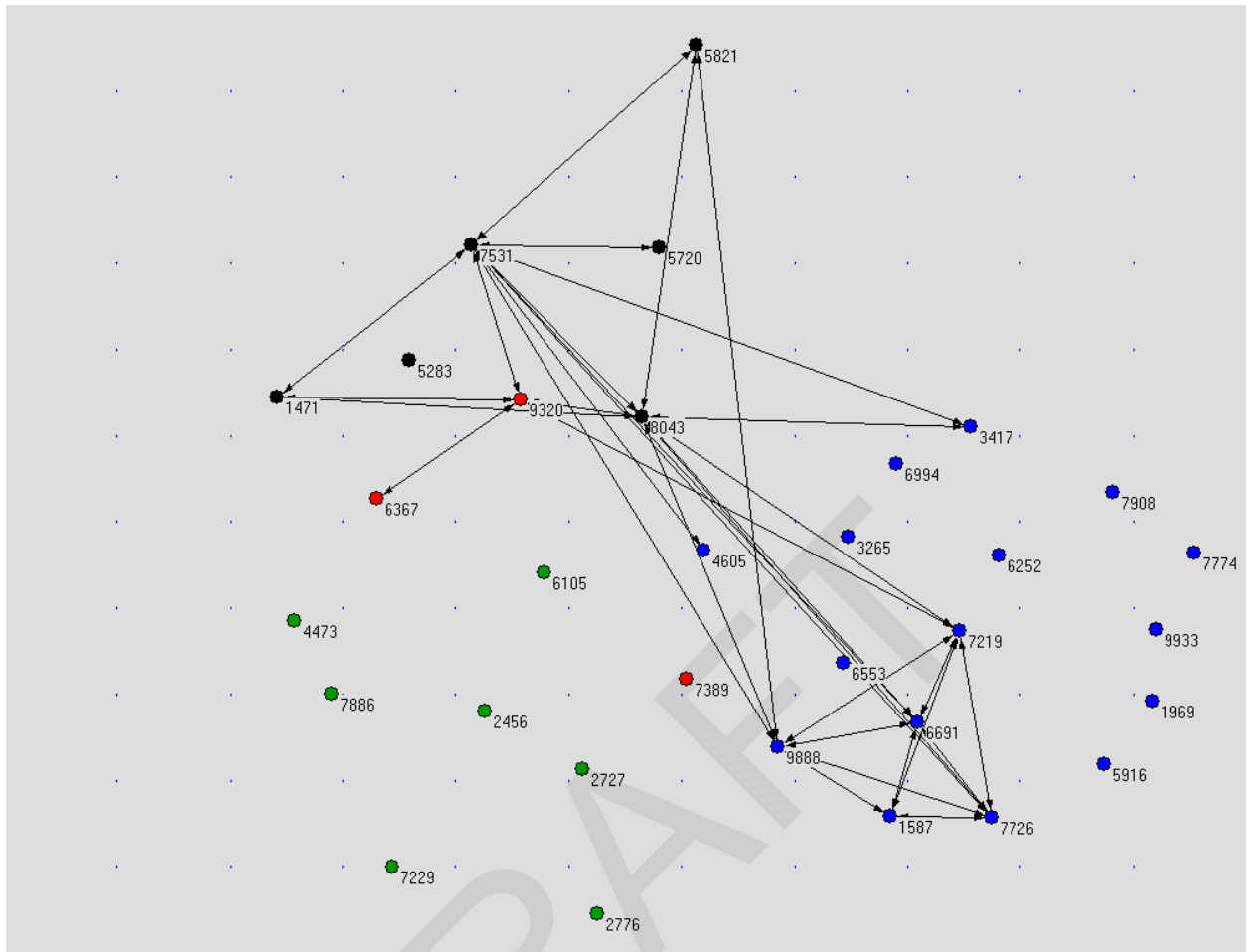
Red = FRIB Leadership; Blue = Engineering; Green = Scientists; Black = Business Support

Figure 3b. Day -6
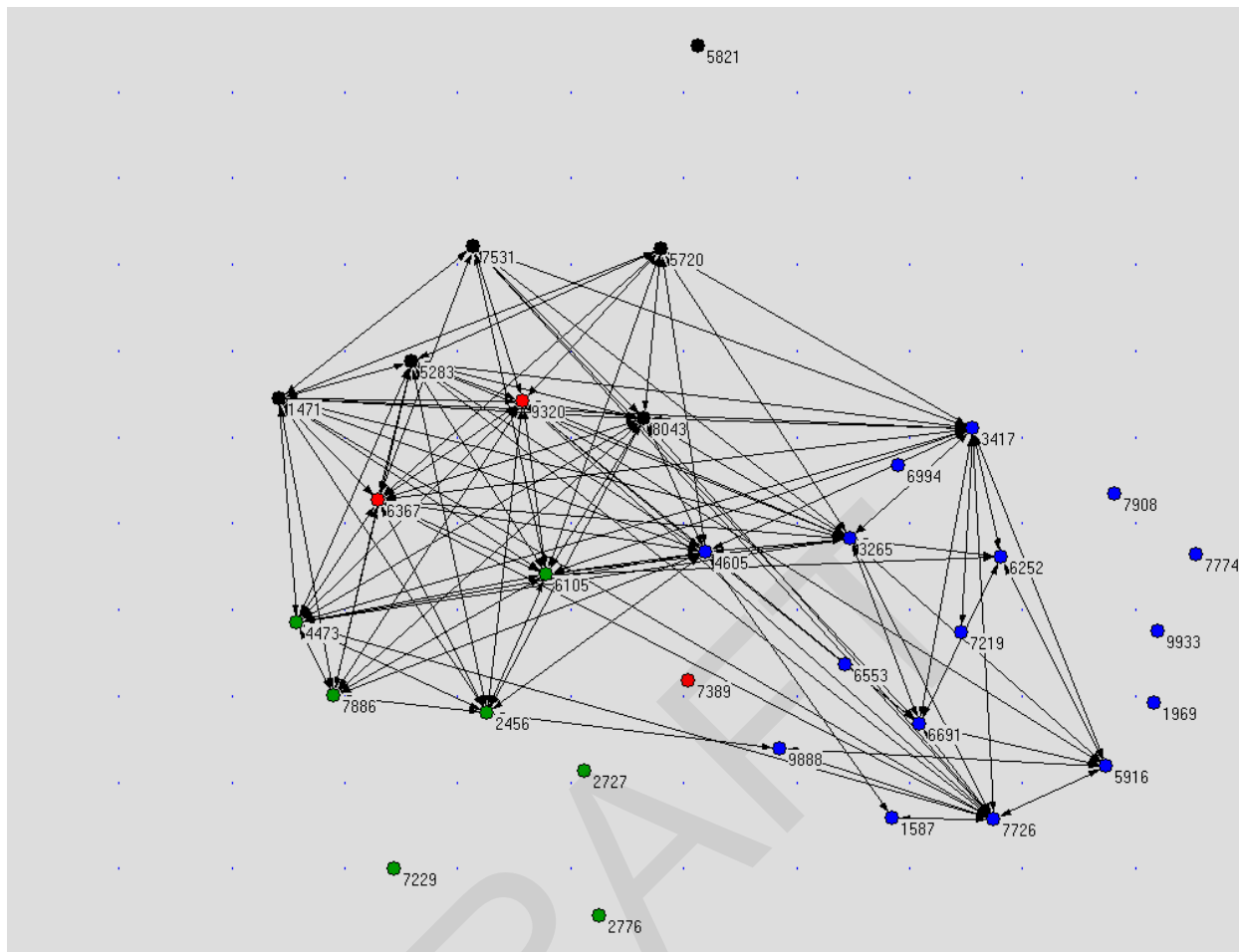
Red = FRIB Leadership; Blue = Engineering; Green = Scientists; Black = Business Support
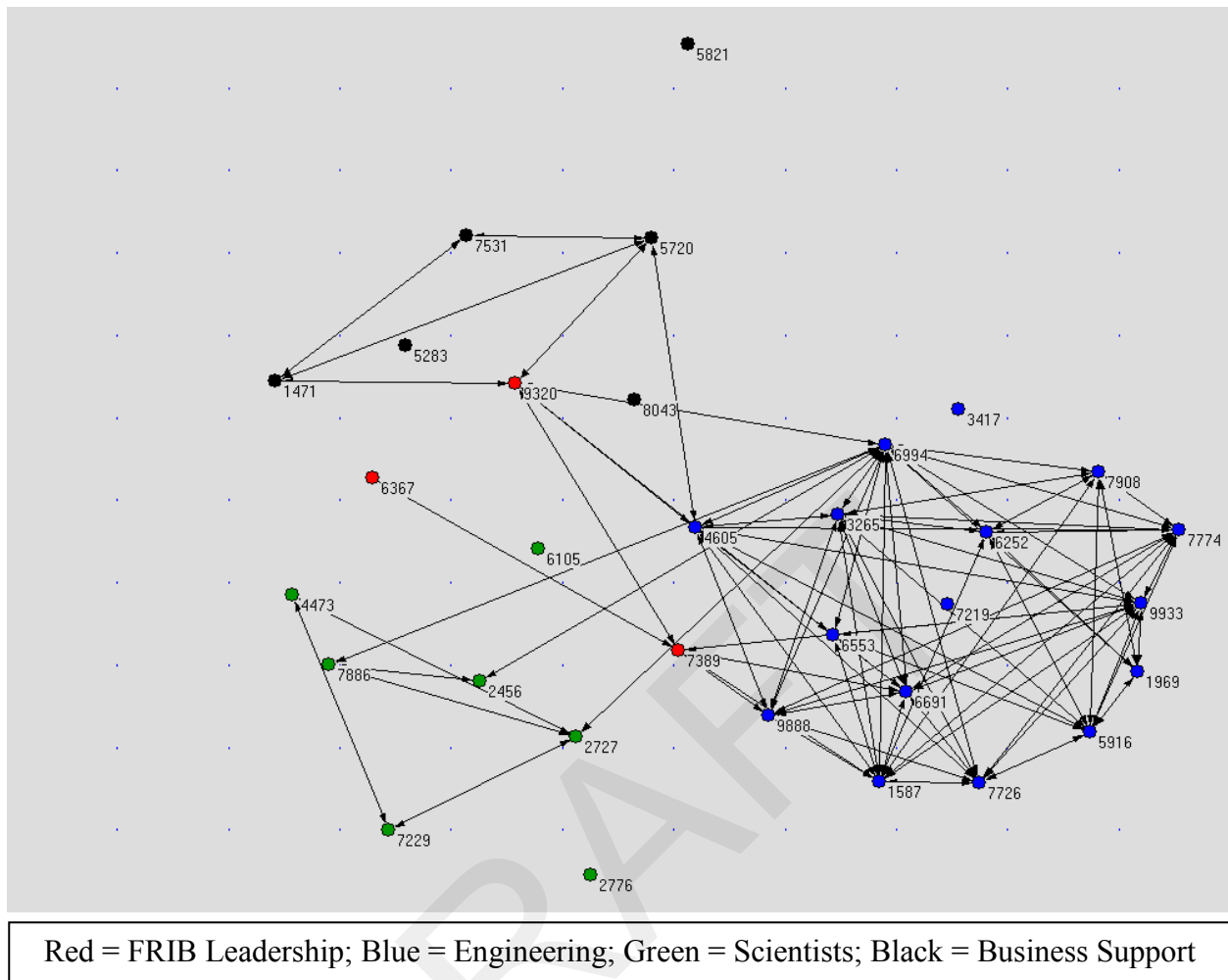
Figure 3c. Day -4
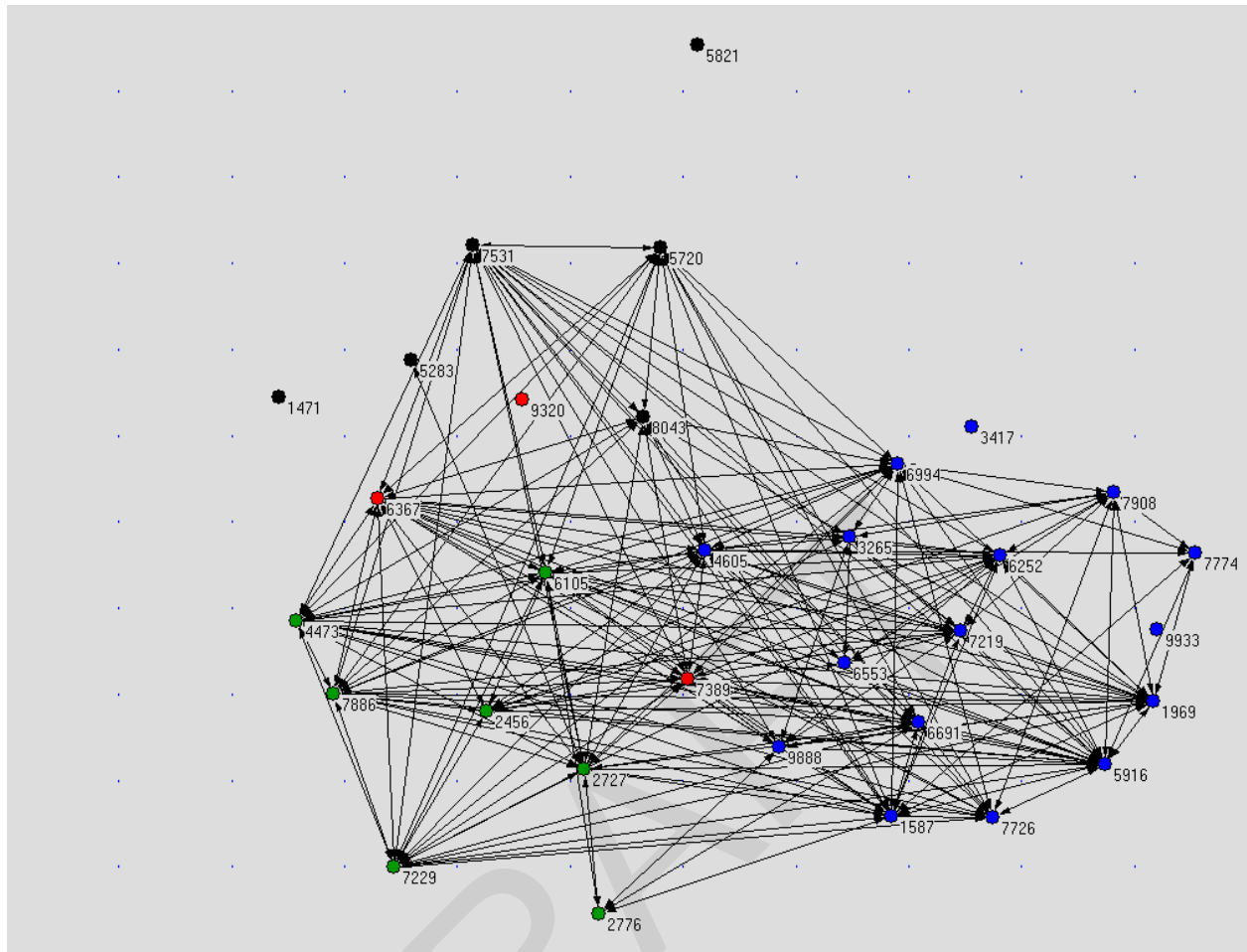
Red = FRIB Leadership; Blue = Engineering; Green = Scientists; Black = Business Support
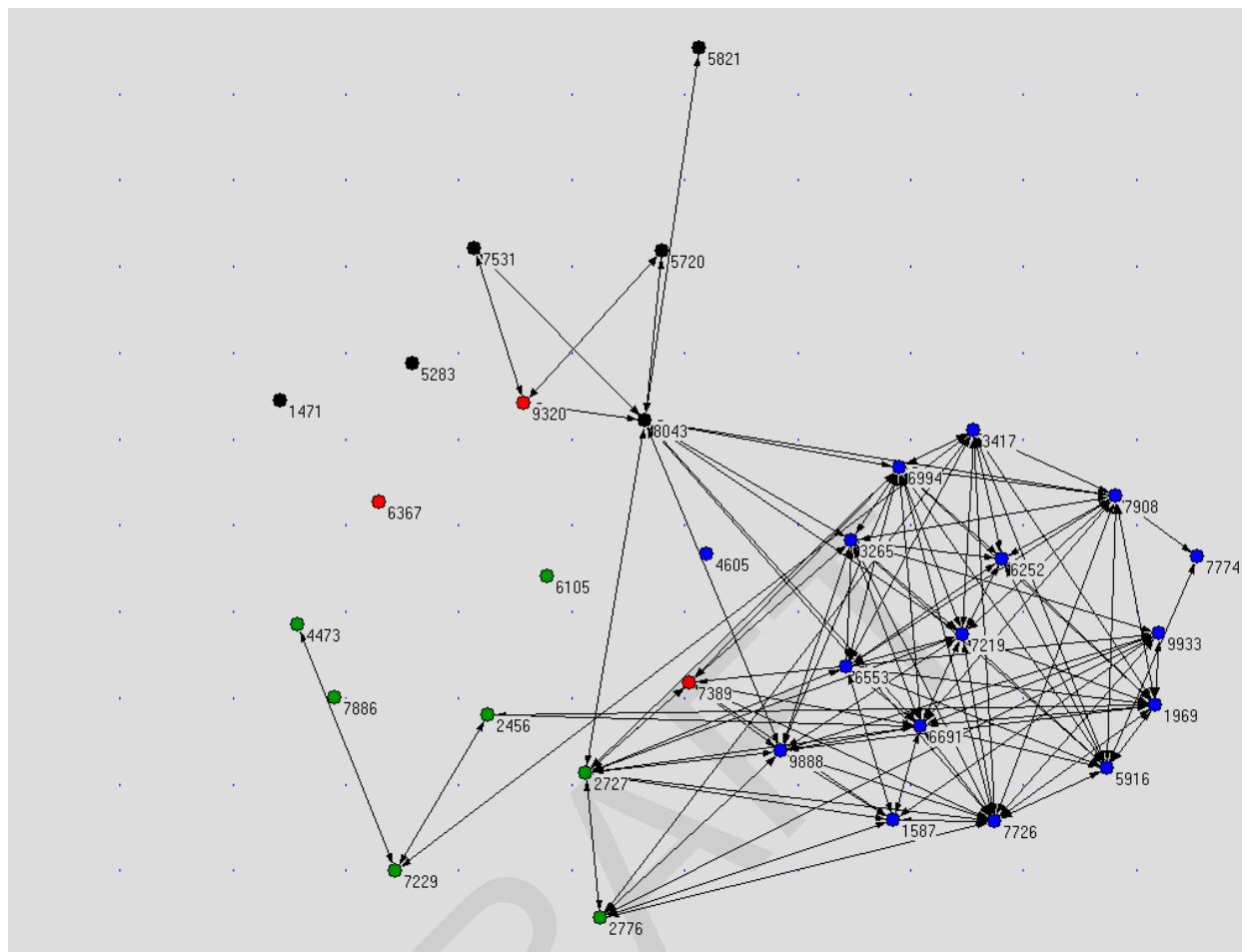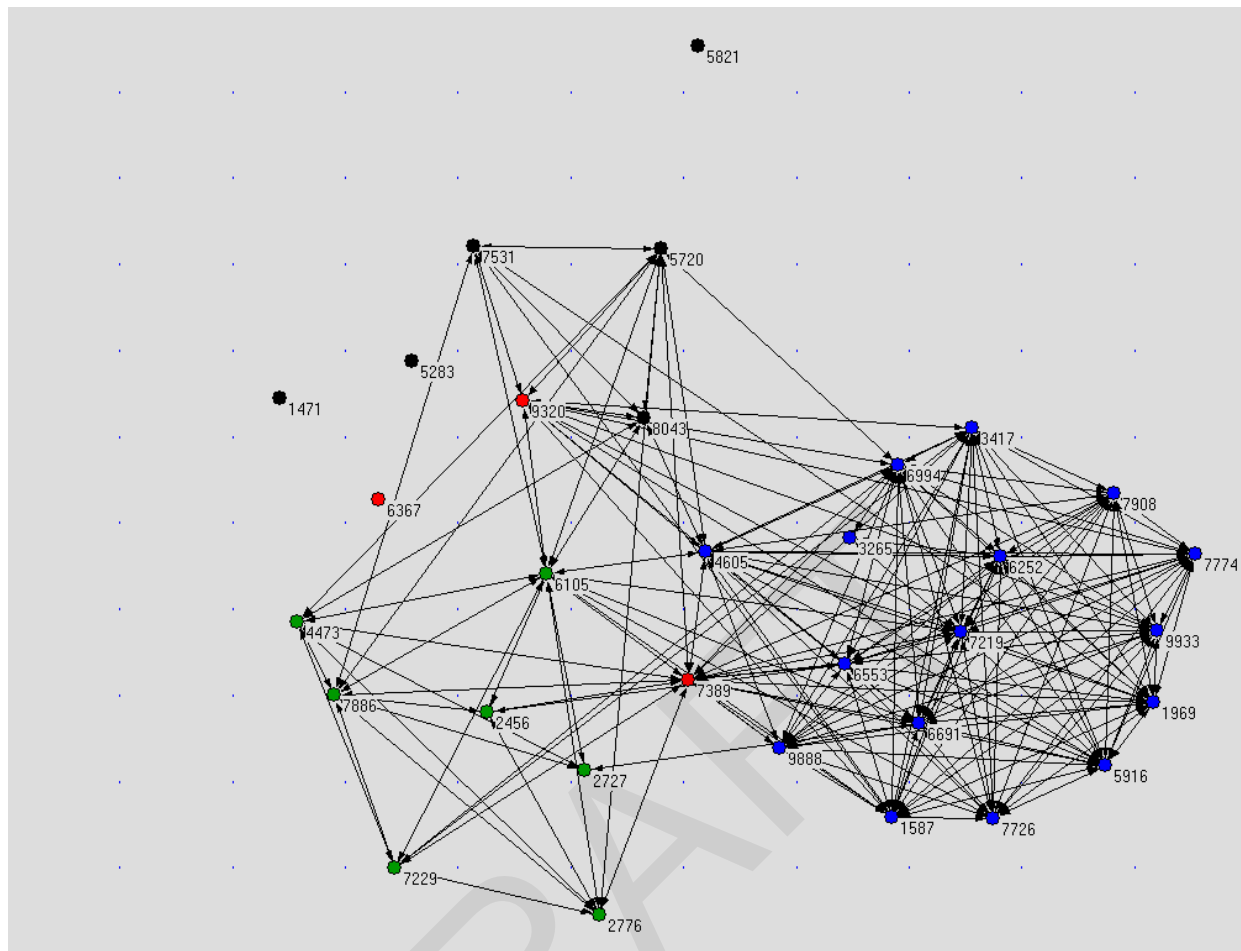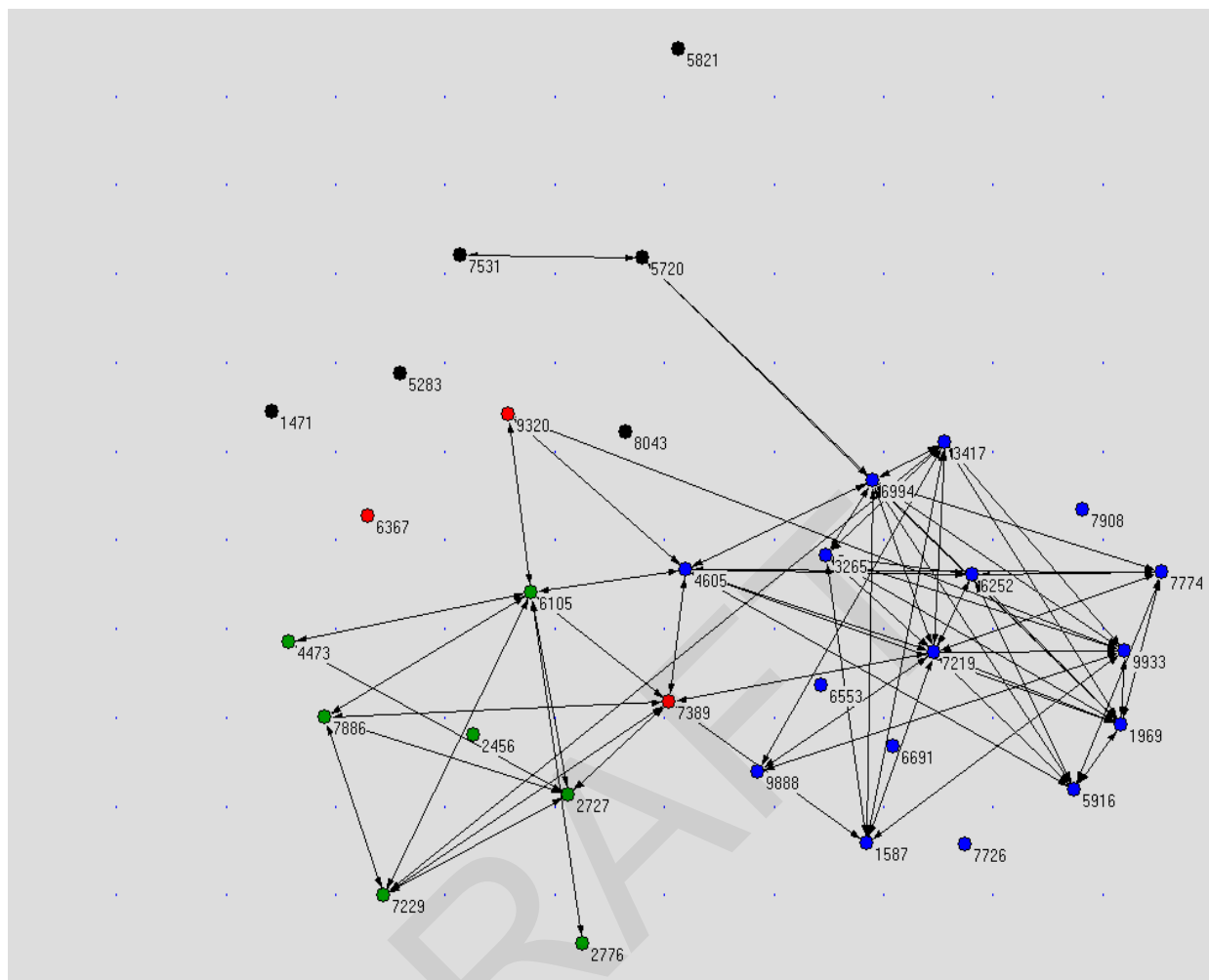
Figure 3d. Day +2
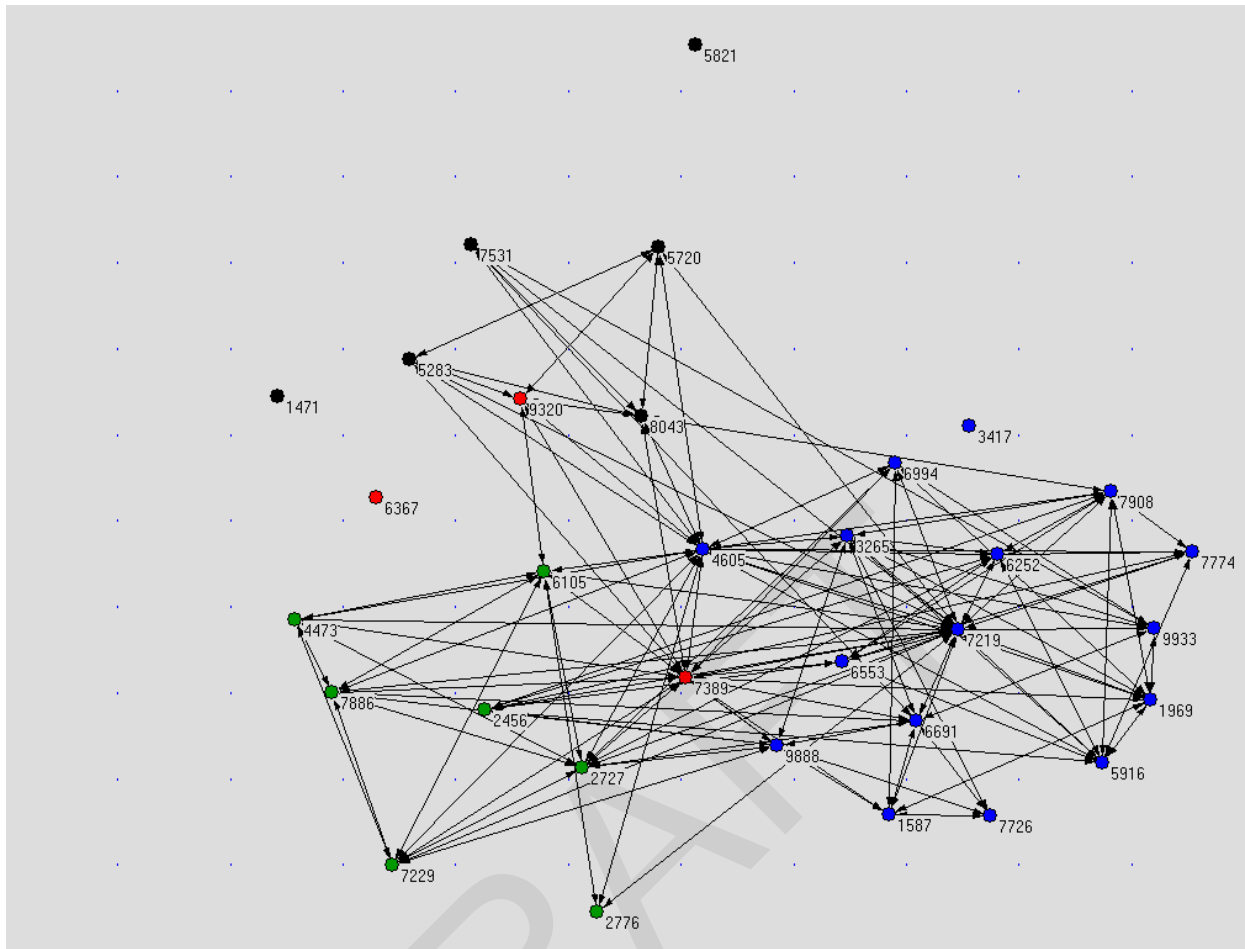
Red = FRIB Leadership; Blue = Engineering; Green = Scientists; Black = Business Support

Figure 3e. Day +4
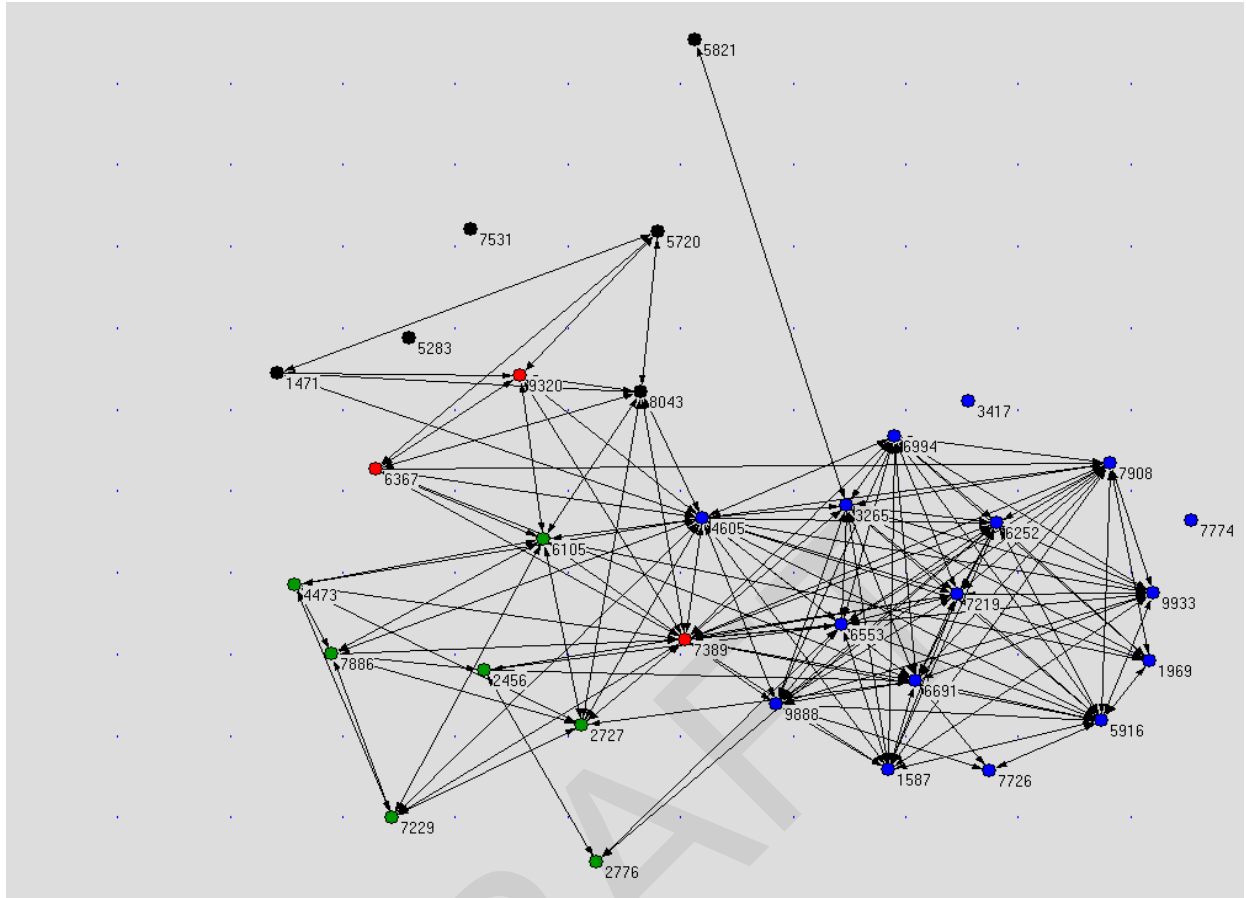
Red = FRIB Leadership; Blue = Engineering; Green = Scientists; Black = Business Support

Figure 3f. Day +8

Red = FRIB Leadership; Blue = Engineering; Green = Scientists; Black = Business Support

Figure 3g. Day +10

Red = FRIB Leadership; Blue = Engineering; Green = Scientists; Black = Business Support

Figure 3h. Day +14

Red = FRIB Leadership; Blue = Engineering; Green = Scientists; Black = Business Support

Figure 3i. Day +16

Red = FRIB Leadership; Blue = Engineering; Green = Scientists; Black = Business Support
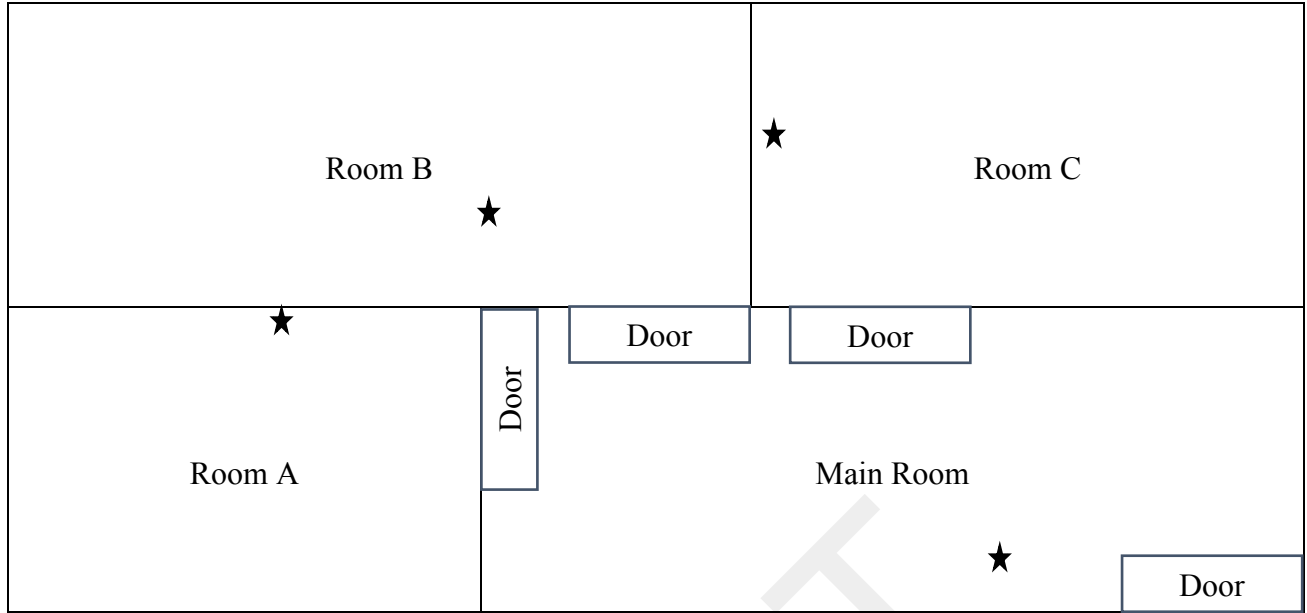
Figure 3j. Day +18

Figure 4. Layout of laboratory suite used in Study 2 with stars indicating base stations