Would you like to be my Facebook friend?
Predicting the formation of Facebook friendships through network overlap, homogeneity and brokerage.
People, whether they be employees of organisations, pupils at schools, athletes in sport clubs, spouses in marriage or siblings in families are members of many, even virtual, social networks. Through these relationships social capital becomes accessible. Social capital is considered to be “the sum of the resources, actual or virtual, that accrue to an individual or group by virtue of possessing a durable network of more or less institutionalised relationships of mutual acquaintances and recognition” (Bourdieu and Wacquant, 1992, p. 119). These resources vary with the type of relationship concerned. One’s closest relationships, strong ties, provide a sense of belonging and emotional support and are associated with subjective well-being and physical health (Helliwell & Putnam, 2004; House, Landis & Umberson, 1988). In contrast, weak ties outside one’s inner circle are particularly important in providing useful information, for example on employment opportunities (Granovetter, 1973).

Putnam (2000) makes a distinction between bonding and bridging social capital. Bonding social capital can be found in tightly-knit, often homogeneous groups and excludes ‘outsiders’. It brings similar people together: “more blacks and more whites to church, but not to the same church” (Putnam, 2000, p.362). Bonding social capital transmits trust and support and is mainly found in networks of strong ties. In contrast, the more inclusive bridging social capital crosses social boundaries. Bridging social capital is found in weak ties, where it facilitates information diffusion and gives access to external resources (Burt, 2007). Both types of social capital are necessary for human flourishing, though from an economical perspective, bridging social capital and networking across boundaries is more likely to ‘get one ahead’ (Burt, 2007; Putnam, 2000). For example, training senior executives to spot opportunities for bridging social capital increased their chances of receiving an outstanding performance evaluation or promotion by nearly 50% (Burt & Ronchi, 2007). But social capital might not be equally accessible to all and can be dependent on personal characteristics.
To research who has access to the different types of social capital it is essential to study the structure of social networks. Although the study of social networks has matured over the last decades (Nohria, 1998) it could still benefit from integration into other mainstream disciplines. The literature on social networks has largely been highly mathematical and technical (Wasserman & Faust, 2008), limiting its application to those familiar with this language (Scott, 1991). In addition, it often underestimates the individual as a passive node. In contrast, I argue that it is necessary to take an individualised approach to the study of social networks by viewing the actor as an individual who actively transforms the structure of his or her social network, instead of a passive node who merely inhabits a place in a pre-formed network of ties. Social psychology could add value to the existing literature by explaining the structure of social networks in terms of the attributes of the individual actors in the models, depicted as nodes (Mehra, 2001). This particular focus on social networks is able to visualise who has access to the resources provided by social capital. A particularly attractive visualisation of the social network, access to social capital and the presence of strong and weak ties is the triad – a group of three actors, who may or may not be tied to each other.

**Triads within social networks**

The triad has gradually established its place in the study of social networks. Already in 1973 Granovetter (p. 1) emphasised the need for research relating “micro-level interactions to macro-level patterns” in his landmark paper on the strength of ties. However, knowledge of the relationship between networks and dyads remains limited (Louch, 2000). Triads could be considered as the intermediate bridge in network analysis as they interconnect dyads to create network structures (Madhavan, Gnyawaliijn & He, 2004). Triads do take the properties of interpersonal relationships in dyads into account, but they have the additional advantage of
being the smallest structure that also visualises the influence of significant others on the dyad. Simple triads can therefore be considered as the “building blocks” of larger social networks (Scott, 1991). The decomposition of the social network into smaller blocks allows for the study of the individual properties of the three actors and is a convenient mean for mapping out the existence and establishment of ties. Triads simultaneously allow research on the micro-level flow of information and its contribution to the direction of social capital.

Triads can either be transitive or intransitive. In a transitive triad, each individual has a relationship, a tie, to both others, for instance a mother and her two children. Transitive triads are at the base of a dense, interconnected network of often strong ties that enable the flow of information within the network (Kalish & Robins, 2006) and contain an element of trust (Burt, 2007). Transitive triads thus appear analogous to the concept of bonding social capital. In comparison, in open – intransitive – triads, one of the three possible ties is missing as the triad theoretically contains weaker ties: a mother shares a tie with her daughter and her co-worker, but her daughter and the co-worker have not formed a tie. Intransitivity in triads creates “structural holes” in the information flow as information from the daughter needs to pass through the mother to reach the co-worker. As individuals between whom structural holes exist often belong to different social groups, intransitive triads are actually a mean to transmit information across groups (Burt, 2007) and could be considered an indicator of bridging social capital.

Balance theory (Heider, 1946) argues that intransitive triads are likely to be converted to transitive triads as they cause “intra-psychic strain” (Curry & Emerson, 1970) in the two individuals among whom no tie yet exists. To reduce this tension they should form a tie between them, thereby resolving intransitivity. Granovetter (1973) argued that this tendency should be especially powerful when the triad contains at least two strong ties, referring to intransitive triads with two strong ties as “impossible triads” as the structural hole in such a
triad should hypothetically always be bridged. Although balance theory has been postulated as a basic social process, its evidence is mixed (Hummon & Doreian, 2003). It bypasses individual human variability. Inescapably, some individuals are more capable of bridging the intransitivity of triads.

It therefore appears that the transitivity of triads is not just influenced by the strength of the ties that form it, but is complemented by other social processes. The following sections will expand on three processes that are expected to influence the likelihood of transitivity in triads, and thus social network integration. It will include social network overlap as a measure of tie strength and draw on theoretical evidence from the homophily principle and brokerage theory. Figure 1 presents a schematic overview of the triad and the three theoretical backgrounds that are expected to influence its transitivity. Reviewing the literature from these diverse theoretical backgrounds will allow to set the first expectations for upcoming research. The final section will present a methodological review which will include an outlined proposal for the required methods, measures and statistical analysis for executing the research.

Figure 1: Expected influences on the transitivity of triads
Social network overlap

A large overlap in social networks and thus a considerable number of shared friends, significantly increases the chance that two individuals will meet and form a tie among them (Granovetter, 1973). Individuals with many shared friends are likely to belong to similar groups and clubs that have provided them with shared social circles (Kossinets & Watts, 2006). In addition, they arguably live in closer proximity than those who share little friends, thereby further increasing the chances of encountering one another. The proportion of shared friends is generally seen as an indicator of tie strength, with strong ties sharing more friends (Gilbert & Karahalios, 2009). Close couples share a greater network of mutual friends (Milardo, 1982). Even in online settings, network convergence occurs when social relationships deepen (Parks & Floyd, 1996).

In triads, the effect of shared networks on tie formation can be twofold. The first effect would take place through the mutual friend. In triads, at least one mutual friend exists as both individual B and C are connected to their shared friend A. Individual A only has a finite amount of time to spend on social networking with his friends. When A is close friends with both B and C and wishes to spend a significant amount of time with either of them, he can spend his time most effectively when he spends time with both of them simultaneously (Granovetter, 1973). As a result of A’s strong tie with both B and C, they are consequently more likely to become acquainted with one another.

A stronger tie, displayed in a higher proportion of shared friends, between A and B and A and C increases the chance of B and C forming a tie and thus increases the chance of transitivity.

Secondly, a related effect can exist through the shared number of friends among B and C themselves. As mentioned, this increases the chances of them encountering each other
through shared activities and interests (McPherson & Lovin-Smith, 2001), resulting in the formation of a tie independent of the agency of actor A. This effect should increase as the overlap in social networks and thus the proportion of shared number of friends, is greater.

A higher proportion of shared friends between B and C increases the chance of them forming a tie and thus increases the chance of transitivity.

Not only do strong ties share more friends, they are also more similar on a number of characteristics (McPherson & Lovin-Smith, 2001), the homophily principle.

Homophily

“Birds of a feather flock together” – the homophily principle – holds for nearly all kinds of social relationships (Brehm, Miller, Perlman, Campbell, 2002; McPherson, Smith-Lovin, & Cooke, 2001). People are more likely to meet, connect better with and transform an acquaintance into a close friend with others who are similar to them on background, personality and interests (McPherson & Lovin-Smith, 2001). As a result, personal relationships are consistently more easily formed within than across social boundaries (Louch, 2000) and people’s social networks are often relatively homogeneous environments (McPherson et al., 2001). Homogeneity in social networks can be attributed to interpersonal attraction to similar others – choice homophily. This desire to seek out interaction with similar others is generally attributed to the need to validate one’s own world view as well as reduce the chance of interpersonal conflict in relationships (Klohnen & Luo, 2003). In addition to seeking out similar others, people are also more likely to come across others like them in relatively homogeneous environments that exist due to institutional and cultural factors –
induced homophily (McPherson & Lovin-Smith, 1987). Both sources of homophily could have an impact on the transitivity of triads.

Lazersfeld and Merton (1954) distinguished two types of homophily, *status* and *value* homophily. I will use this distinction to categorise a number of personal characteristics that is open to study.

*Status homophily*

Status homophily or demographic homophily includes ascribed characteristics such as ethnicity, gender and age as well as acquired characteristics such as occupation and education. Homophily in ethnicity generally creates the strongest social divide, with age, education, profession and gender gradually becoming less influential (McPherson et al., 2001). Due to restrictions in the method of data collection, only homophily in gender and age will be studied.

- Homophily in gender. Gender is a demographic characteristic which has, in contrast to other individual variables such as race, an equal baseline distribution. Social networks could, therefore, contain equal numbers of men and women. However, networks are influenced both by the personal preferences of its actors as well as societal structures that increase the probability of same-sex contact. Such induced homophily is found in both professional as well as voluntary organisations (Williams, 1992; Popielarz, 1999): glass walls supposedly prevent both genders to join professions that are traditionally male- or female-dominated, whereas glass ceilings mainly exclude women from reaching equal representations in top functions (Guy & Newman, 2004). Gender segregation also occurs by choice outside of work, for instance in sport teams. Such segregation makes the formation of homogeneous ties more likely than chance would suggest. Although gender
has one of the least powerful effects on homophily - less so than age, race and education - considerable gender homophily can be found in non-kin confidant networks (Marsden, 1988), friendship circles (Verbrugge, 1977) and face-to-face groups in voluntary organisations (McPherson & Lovin-Smith, 1987).

- Homophily in age. Age homophily varies significantly with the type of tie studied. Close ties are especially prone to homophily of age, which is usually stronger than nearly all other dimensions. In two samples of participants, Verbrugge (1977) demonstrated that up to 50 percent of one’s closest friends fall into the same age category. In addition, a study by Feld (1982) suggests that age homophily is very common in weaker ties at work and in the neighbourhood as up to 40 percent of friends in these situations are within a range of five years. Age homophily appears substantially lower in kinship networks, where ties are formed across generations. Age homophily is considerably enlarged by institutionalised practices such as the grouping of school classes according to age (McPherson et al., 2001). This stratification continues into adulthood when joining the work force. Except for kinship ties, such practices would increase the chance of age homophily in nearly any tie.

**Value homophily**

Value homophily is based on shared attitudes and beliefs. It includes personality, interests and religious and political convictions. A study by Wetzel and Insko (1982) suggests that people are mainly attracted to others who conform to their ideal self more than their actual self. This should be most rewarding in interaction and adheres to the desire to seek out other’s who confirm one’s world view (Klohnen & Luo, 2003). As attitudes, beliefs and values are closer to one’s ideal self than personality characteristics, the preference to associate with others who
are similar to the self is especially strong for beliefs, values and attitudes (Wetzel & Insko, 1982) and less so for personality (Luo & Klohn, 2005).

The proposed research allows the following indicators of values to be taken into account: personality, religious views, interests and favourite music and movies. The effects of value homophily can appear exaggerated when individuals assume others to be similar to them and avoid discussions of differences (McPherson et al., 2001). The study will therefore focus on actual similarity by comparing the similarity of the self-reported interests, beliefs and personality characteristics instead of relying on direct self-assessed estimates of similarity.

In one of the few studies on the effects of homophily on the transitivity of triads, Louch (2000) found that in triads of strong ties, homophily of nearly all characteristics exerted significant effects on the transitivity of triads. All three actors in a triad were more likely to be connected to each other when they were more similar to each other in age, race, education and religion. Only gender showed a surprising effect as close confidants in these triads were actually less likely to form ties with alters of the same gender. However, this effect was mitigated by marriage: the same-sex friends of a married respondent were less inclined to befriend the respondent’s spouse. When this effect was eliminated from the model, gender homophily became present, although it was more weakly related to transitivity than the other characteristics.

Status and value homogeneity increase the probability of an alter pair forming a tie. As a result, individuals B and C, in the ‘missing’ dyad are more dissimilar than actors who share a tie.
**Brokerage**

Burt (2007) distinguishes two types of roles: the broker and the closer. Brokers are only found in intransitive triads. They are the intermediate actor in an intransitive triad: connected to both alters who have not formed a tie among themselves. The closer has closed this structural hole, creating a transitive triad. The broker is thus in a relative position of bridging capital, whereas the closer has better access to bonding capital. Relatively little is known about the personal preferences of individuals for the role of broker or closer. However, brokerage appears to be associated with both demographic and personality characteristics.

Traditional brokerage models argue that the role of broker allows for the better pursuit of one’s own individual interests and is viewed as a position of power (Nohria, 1998). This is arguably a more masculine function. Women’s networks appear to be denser, more woven into the ‘social fabric’ (Campbell & Barrett, 1992) and contain more strong, kinship ties (Moore, 1990) that would arguably serve a social support function (Burt, 1998). In professional organisations women are even more likely to reap benefits, such as a promotion, from a small network of interconnected ties (Burt, 1998). Women can thus be expected to be more attracted to and fare better in transitive triads, whereas men are assumed engage more frequently in and benefit more from intransitive triads.

In addition to gender, brokerage also appears to be associated with personality. In a study by Burt, Jannotta and Mahoney (1998) MBA-students whose networks bridged structural holes were more likely to have an ‘entrepreneurial’ personality – an outsider in search of change and authority opposed to those who were more comfortable in a closer role. Another study by Kalish and Robins (2006) showed that brokers with two strong ties to both alters were more likely to have an individual focus, whereas extraversion was positively associated with closed triads of strong friends and neuroticism with closed triads of weak ties. The rationale behind this is that gregarious extraverts inhabit larger social networks in which
they are more likely to introduce their friends to each other, thereby fulfilling the role of broker. Surprisingly, Kalish and Robins found no effect for self-monitoring. As high self-monitors observe themselves and tailor their behaviour to the needs of the social situation (Snyder, 1974), they should be able to get alone with a diverse spectrum of people which should provide them with opportunities of bridging structural holes and maintaining a broker position.

To research the relationship between brokerage and personal characteristics, the proposed study will include the Big Five personality characteristics and self-monitoring as measures of personality and gender and age as measures of demographic characteristics.

*Brokerage is associated with certain demographic characteristics (male) and personality characteristics (*i.e* extraversion).*

**Methods**

A relatively new social medium has been chosen for the collection of data: Facebook. Recent years have seen a surge in the amount of studies on Facebook and other social network sites (SNS’s)\(^1\). These websites are a particularly attractive medium for the study of social networks. Social networking sites such as Facebook present strong ties with close friends, but are especially important as a low threshold method for activating latent and maintaining weak ties. Facebook is possibly the largest of all SNS’s, claiming to have over 500 million active users, with every user tied to on average 130 ‘friends’ and over 70 percent of users coming from countries other than the United States\(^2\). Unlike some other SNS’s, these friendships

---

\(^1\) The definition of Social Networking Site (SNS) that is used in this context is taken from Boyd & Ellison (2008): web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system.

appear to reflect off-line ties (Clouston, Verdery, Amin & Gauthier, 2009): used to maintain or strengthen existing offline relationships rather than for meeting strangers (Ellison, Steinfield & Lampe, 2007). In a study by Mayer and Puller (2008) students reported that only 0.4% of their Facebook friendships represented an exclusively online relationship. Arguably, most of the average 130 Facebook friendships would be weak ties, but the friendships are likely to have been formed around a class, job, sport club or holiday together. Facebook thus provides a source of naturalistic behaviour in a virtual setting (Boyd & Ellison, 2008). For the study of Facebook friendships it can therefore be assumed that similar social processes are at work in this virtual social network as in the offline world which will allow this study to draw on literature of offline social interactions. The study of online social networks has the additional advantage that it does not impose boundaries on the size of the participant’s networks, as in classic network studies, removing a possible bias (Lewis, Kaufman, Gonzalez, Wimmer & Christakis, 2008). Furthermore, the collection of data which can be mined practically automatically from a large and geographically widespread number of participants proves relatively easy.

Facebook’s popularity and accessibility have made it a breeding ground for an enormous number of applications that allow users to play online games, read their daily destiny through fortune cookies and rate the hotness of complete strangers. The data for the proposed study have become available through the more serious MyPersonality Application, which allows Facebook users to take scientifically validated psychometric tests. The application currently allows users to take several of the more than a dozen questionnaires available with more questionnaires able to be uploaded at any time. It is expected that users are highly motivated to answer test questions to the best of their knowledge as the self-selection procedure assumes a certain curiosity into one’s own personality which will only be fulfilled if questions are answered truthfully. Although the outcome of each questionnaire can
be published on user’s Facebook profile page, most users do not share their results which should further reduce the influence of impression management.

When adding the application to one’s Facebook profile, users are asked for access to their personal profile data. These data, along with the results of any questionnaire taken, are stored in the MyPersonality database at the Psychometrics Centre of the University of Cambridge. The database currently contains data from nearly 5 million users, of which around 40 per cent have granted access to their full personal profile data. Personal profile data include: age, gender, religious and political views, interests, favourite movies, music and books and a generated list of friends. All respondents have been anonymised through identity numbers.

Although the ease of access and sheer numbers of data collection through the MyPersonality database are very attractive, a critical note must be added. Firstly, the database uses convenience sampling, a less rigorous method of participant recruitment than probability sampling (de Leeuw, Hox, & Dillman, 2008). However, there is no immediate reason to suspect that millions of users who have selected themselves for personality questionnaires behave significantly different in social networks than those who have not. Furthermore, among younger generations in several Western countries Facebook penetration has reached up to 40 percent of the population, making it likely that a diverse sample of the population is represented. Nevertheless, sampling bias will be checked for by comparing the demographic characteristics of the MyPersonality database against the general Facebook population.

---

1 More information about the questionnaires and current research available through the application can be found at the MyPersonality Wiki on http://www.mypersonality.org/wiki/doku.php?id=available_data. The MyPersonality application can be accessed through a Facebook account on http://www.facebook.com/apps/application.php?id=2490151219&v=info.

Ultimately, convenience sampling will warrant some caution when generalising the results of the study to other populations.

A second, more substantial obstacle is the limited control over the collection of personal data. Data collection is firstly restricted to the data that Facebook requests of its users. For instance, race will not be included in the current study as Facebook does not ask this of its users. Furthermore, the self-report options of some Facebook profile variables provide full freedom, resulting in “humorous” statements on religious and political views ranging from “manbearpig”, “lipstick” to “Oprah”. This makes these responses incredibly difficult to code. Religious and political views will therefore have to be excluded from the analysis. Favourite movies, music and books on Facebook are still able to be included as these are displayed by matching a user’s input to an existing artist, movie or book, which generates a more standardised list of self-reported information.

Respondents
From the MyPersonality database, a total of unique 3238150 triads in which one individual shares a tie with at least two alters in the database have been computed. All triads thus contain three individuals who share at least two ties. Individuals can be included in multiple triads. The age distribution of the data base is positively skewed with a relatively young user sample ($m= 22.97$ years; $sd=6.93$). The minimum age of respondents is 18, as Facebook requires users to be over 18 for the use of this application, with no respondents in the triad database over 60. Women are slightly overrepresented, as 54 percent of respondents are female.

Measures
Transitivity has been computed by examining the existence of Facebook friendships between three actors in the MyPersonality database. Triads in which only two alter pairs share a
friendship are marked as intransitive; triads in which all three actors share a tie are labeled as transitive. The great majority of triads, 94.4%, are intransitive, indicating that only in one out of 20 triads all three actors share a tie. This is considerably lower than previous research on strong ties (Louch, 2000: 21% intransitive triads) and could be attributed to the relative weak, geographically widespread ties that exist on Facebook.

Personal profile data on age, gender, interests, favourite movies, music and books and total number of friends have been extracted from the MyPersonality database for all three individual actors. Gender is assigned 0 for female and 1 for male; age is calculated from the birthday on the Facebook profile and is reported in years. Interests, favourite movies, music and books have been copied from the self-reported sections on Facebook. The total number of friends has been calculated from a list of user’s friends. By comparing individual’s data to their alter’s, homogeneity of each dyad on the included characteristics has been computed.

Questionnaires for personality variables firstly include the International Personality Item Pool Representation (IPIP) of the NEO PI-R (Costa & McCrae, 1992). The NEO PI-R is divided in five personality factors with scores ranging from 0 (low) to 5 (high) on each factor: openness ($m=3.85; sd=.72$), conscientiousness ($m=3.39; sd=.78$), extraversion ($m=3.75; sd=.81$), agreeableness ($m=3.49; sd=.76$) and neuroticism ($m=2.66; sd=.83$). This five factor model of personality characteristics is currently the most widely used model for the structure of personality (Costa, 1996). The triad database contains the full test scores on all five personality factors of 2899267 users. Secondly, Snyder’s (1974) 25 item self-monitoring questionnaire contains relevant questions on the extent that individuals monitor and adapt their behaviour to social situations. It has gathered 32354 test scores in the triad database ($m=13.31; sd = 4.05$).

Network overlap has been computed from the number of shared friendships. An overlap in friendship circles is generally considered to be a strong indicator of tie strength.
(Granovetter, 1973) and has also been shown to be one of the most important predictors of Facebook tie strength (Gilbert & Karahalios, 2009). As the number of shared friendships is dependent on the total number of Facebook friendships of both actors in a dyad, the strength of a Facebook tie has been calculated from shared friendships as a proportion of the average total number of friends of the alter pair. Network overlap appears to differ considerably for transitive and intransitive triads. Transitive triads are considerably denser than intransitive triads with the average number of mutual friends in transitive triads 66 ($sd=70.25$), which amounts to a proportion of 0.10 of all Facebook friends ($sd=0.08$). Intransitive triads, however, display considerably weaker ties: friendships between the broker and the two alters contain an average of 35 mutual friends ($sd=51.44$), which equals to four percent of all their friends. The ‘missing’ friendship in intransitive triads unsurprisingly displays an even weaker pattern with an average of 9 mutual friends ($sd=21.42$) between the individuals who do not share a tie – a proportion of 1 percent of their total number of friends ($sd=0.03$).

**Statistical Analyses**

To predict the transitivity of a triad, a dichotomous outcome with 0 as intransitive and 1 as transitive, logistic regression can be used with at least one continuous variable included in the model. When studying the effect of homogeneity on transitivity, the influence of shared friends will be controlled for in the logistic regression. In addition, to test whether individuals who inhabit the broker position are significantly different from their non-broker counterparts, independent t-tests will be performed.

As the dataset contains very large numbers of triads, it is likely that some analyses will show statistical significance ($p<.05$), but might lack practical relevance. Apart from reporting the significance of any test, it is therefore critical to take effect size into consideration and discuss its implications compared to significance values.
The proposed study will cover the literature and methods more deeply, provide results on the influence of social network overlap, homogeneity and brokerage on transitivity in Facebook triads and critically discuss the implications and limitations of the research. I believe that this research will be able to make an original contribution to knowledge on the influences of friendship formation and the study of social networks in a naturalistic, virtual, setting as well as test the theoretical background using a unique database with unrivalled numbers of diverse participants.
References


