Comparing Degree-based and Closeness-based Centrality Measures

Christian Tallberg*

Abstract

Statistical properties of four centrality measures are investigated. Three of the measures are extensively used within the context of social networks. The measures are investigated for two centrality concepts, degree centrality and closeness centrality. The model assumed to generate realizations of social networks is a conditional Bernoulli model where the edge probabilities are independently beta distributed. For various shape parameter values of the beta distribution, properties of the graph centrality measures are compared for degree-based centrality and closeness-based centrality.

Three of the graph centrality measures exhibit similar behaviour when they are degree-based and closeness-based. The fourth measure, the variance of the actor centralities, generate diverging tendencies of the properties of the measure depending on whether it is degree-based or closeness-based. This is interpreted as an indication that the measure captures different features of the structures of the social network when it is degree-based and closeness-based.

Keywords: Degree centrality, Closeness centrality, Bernoulli model, Beta distribution

*Department of Statistics, Stockholm University. E-mail: Christian.Tallberg@stat.su.se. The author would like to thank Ove Frank for helpful comments. Partial financial support from the Swedish Council of Research in Humanities and Social Sciences (HSFR), grant No. F0750/96, is gratefully acknowledged.

Contents

1 Introduction

In various connections it is of interest to investigate properties of structures. Social network analysis for instance is devoted to the structure of social groups. One important property of such groups involves the degree of importance or popularity of the members of it. The term, frequently used in the literature for this concept, is centrality.

When centrality is based on degree, statistical properties such as estimators of parameters or probability distributions of estimators are relatively easy to derive analytically. For example, if the edges in an undirected graph are generated independently from a Bernoulli(p)-distribution, the probability distribution of the variance of the degrees is shown to be satisfactorily approximated with gamma distributions according to Hagberg (2000). Another example is given by Frank (2000) who considers a directed graph where the edges from i to j are assumed to be Bernoulli (p_i)-distributed for $i \neq j$. If the edge probabilities p_i are beta (α, β)-distributed, the relative out-degree is shown to be an unbiased estimator of the mean of p_i. To perform similar analytical derivations when centrality measures are based on closeness is difficult. Therefore one has to rely on results obtained by computer simulations. The focus here is to investigate relationships between properties of degree-based and closeness-based graph centrality measures from different aspects. Based on the outcome of the simulation results, attempts are then made to formulate statistical properties when centrality is closeness-based. The results obtained by computer simulations are compared to well-known theoretical derivations of statistical properties when centrality is based on degree. For a specific sufficiently large number of simulations the results based on degree agree with the theoretical. It is assumed that this number of simulations is also large enough to assure that the results agree with the theoretical when centrality is closeness-based.

Notation needed is given in the next section. In Section 3 distributions of various centrality measures are discussed and compared. In Section 4 two approaches are taken to compare degree-based centralities with closeness-based centralities. In Section 5 some associations between graph centrality measures are investigated.

2 Notation

Let G be a directed random graph of known order n according to Frank (2000). By convention no loops are allowed. The structure of G is specified by its corresponding $n \times n$ adjacency matrix A where the elements are given by

$$a_{ij} = \begin{cases} 1 & \text{if there exists an edge between vertex } i \text{ and vertex } j \\ 0 & \text{otherwise.} \end{cases}$$

The elements a_{ij} are conditionally independent Bernoulli(p_i)-distributed, where p_i are independent and assumed to be beta (α, β)-distributed. The density of a beta distributed variable is given by

$$f\left(x \left| \alpha, \beta \right.\right) = \frac{1}{B\left(\alpha, \beta\right)} x^{\alpha-1} \left(1-x\right)^{\beta-1}, 0 \leqslant x \leqslant 1,$$

where $\alpha > 0$ and $\beta > 0$ are shape parameters and $B\left(\alpha, \beta\right)$ denotes the beta function,

$$B\left(\alpha, \beta\right) = \int_0^1 x^{\alpha-1} \left(1-x\right)^{\beta-1} dx.$$

The outdegree of actor i, the number of vertices adjacent *from* actor i, is denoted a_i and given by the row sums of the matrix A

$$a_i = \sum a_{ij}.$$

The sum of the reciprocal distances *from* actor i to all other actors is denoted c_i and given by

$$c_i = \sum_{j \neq i} \frac{1}{d_{ij}},$$

where the element d_{ij} is the length of the geodesic from vertex i to vertex j. If no geodesic exist from vertex i to vertex j, i.e. the vertices are located in different components, then $d_{ij} = \infty$ and $1/d_{ij}$ is defined to equal zero. For further conceptual clarifications the reader is referred to Wasserman and Faust (1994).

One of the primary issues within the field of social networks is to investigate the structure of the interrelations of the actors. An important aspect of the structural properties of the actors is centrality. There are mainly three concepts of centrality: degree, closeness and betweenness. Freeman (1979) gives a detailed discussion of actor centrality and network centrality. If the actor centrality indices are aggregated across all actors we obtain a measure of centrality on group level, graph centrality. Some writers advocate the average of the actor centralities as a measure of graph centrality. Bavelas (1950) and Sabidussi (1966) reasoned that a group centrality measure indicates the compactness of a graph and proposed group measures based on the inverse of sums of actor indices. A more common approach to graph centrality is that a group level index should reflect the heterogeneity of actor centrality indices, see Nieminen (1974) and Freeman (1977). A graph heterogeneity measure suggested by Snijders (1981a, 1981b) and also considered by Hagberg (2000) is the variance of the degrees. Two alternative measures of graph centrality are discussed in Tallberg (2000), the maximum of the actor centrality indices, and the difference of the maximum and the average of the actor centralities. In the context of centrality in social networks, the maximum of the centralities, the average of the centralities and the variance of the centralities are common and interesting graph centrality measures. Statistical properties of these three measures and the difference of the maximum and the average of the actor centralities are discussed for degree-based centrality and closeness-based centrality.

In this study we consider four degree-based centrality measures:

- $T_1 = \max_i a_i$
- $T_2 = \bar{a} = \frac{1}{n}\sum_i^n a_i$
- $T_3 = \max_i a_i - \bar{a}$
- $T_4 = s_a^2 = \frac{1}{n}\sum_i (a_i - \bar{a})^2$

and the following corresponding four closeness-based centrality measures:

- $T_1' = \max_i c_i$
- $T_2' = \bar{c} = \frac{1}{n}\sum_i^n c_i$
- $T_3' = \max_i c_i - \bar{c}$
- $T_4' = s_c^2 = \frac{1}{n}\sum_i (c_i - \bar{c})^2$

3 Simulated distributions

When investigating statistical properties of estimators in random graphs, a simple model frequently assumed to generate graphs is the Bernoulli (p)-distribution with a fixed edge probability p. Hagberg (2000) gives a detailed

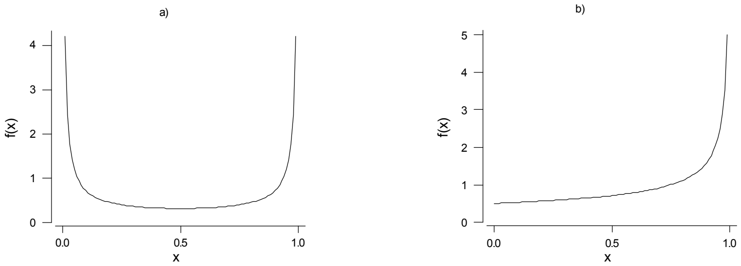

Figure 1: Densities of beta (α, β)-distributed variables; a) beta $(0.2, 0.2)$-distribution; b) beta $(1.0, 0.5)$-distibution.

discussion about properties of the variance of the degrees under this model. The same model is used as a null hypothesis model by Karlberg (1999) for testing transitivity in digraphs, and by Tallberg (2000) for testing centrality in random graphs. A more realistic approach would be to assume that not all the edge probabilities are equal in a random graph. Frank (2000) gives an example where he suggests that the edge probabilities are latent centrality properties, which are assumed to be independently beta (α, β)-distributed.

Guided by this idea we focus on relationships between properties of degree-based quantities and closeness-based quantities for various values of α and β. In this section simulated distributions of the degree-based measures and the distance-based measures are compared.

As a simple example, the similarity of distributions of degree-based quantities and closeness-based quantities are shown when p_i are independently beta (α, β)-distributed for $\alpha = \beta = 0.2$ (Figure 1a)) and $\alpha = 1.0, \beta = 0.5$ (Figure 1b)). The probability distributions are estimated by computer simulations based on 10000 replications (simulated graphs) for the eight quantities mentioned in Section 2.

In non-directed random graphs where the edge probabilities are assumed to be equal between all actors, the distributions of some of the quantities can be derived. For example, the mean degree is $(2/n)\text{bin}\left(\binom{n}{2}, p\right)$-distributed. Furthermore, Hagberg (2000) gives a detailed discussion about how to approximate the distribution of the variance of the degrees with a gamma(θ_1, θ_2)-distribution in the following way:

Consider a sequence X_i of n independent identically distributed normal random variables with mean μ and variance σ^2. Let $s^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$ where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. In a Bernoulli graph the degrees of the vertices are binomially distributed with $\mu = (n-1)\,p$ and $\sigma^2 = (n-1)\,p\,(1-p)$. By making use of that binomially distributed random variables are approximately normally distributed he argues that the variance of the degrees is approximately gamma $((n-1)/2, 2(n-1)p(1-p))$-distributed. Due to dependence between the vertex degrees he suggests a correction that improves the approximation of the gamma-distribution.

If we instead would consider a directed random graph where p_i are assumed to be independently beta(α, β)-distributed, the distribution of the mean outdegree conditional on the outcomes of p_i is $(1/n)\text{bin}\left(2\binom{n}{2}, p\right)$-distributed. Implementing the result of Hagberg (2000) to directed random graphs conditional on p_i yields that the variance of the degrees is approximately gamma $((n-1)/2, 2(n-1)p(1-p))$-distributed. Since there is independence between vertex out-degrees, adjustment is not required.

The unconditional distribution is known for T_2, but not for the other measures. If we let $y = (n-1)T_2$, then according to well-known properties of the beta distribution and the binomial distribution the joint distribution of y and p is given by

$$f(y,p) = \binom{n-1}{y} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{y+\alpha-1} (1-p)^{n-y+\beta-2}.$$

The marginal probability distribution of Y is

$$f(y) = \binom{n-1}{y} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(y+\alpha)\Gamma(n-1-y+\beta)}{\Gamma(n-1+\alpha+\beta)},$$

a distribution known as the beta-binomial.

To derive the distribution of s^2 analytically should be difficult since s^2 is gamma-distributed, and the beta family is not conjugate for the gamma family.

Note the resemblance of the distributions of the degree-based measures and the closeness-based measures, according to Figures 2 and 3. An exception is perhaps the variance of the centralities. The distributions of T_2, T_3, T_4, T_2', T_3' and T_4' should be satisfactory approximated with a normal(μ, σ)-distributions for sufficiently large n.

4 The influence of edge probabilities on centrality

4.1 Some results on moments of degree centrality

Consider random graphs where the edge probabilities, p_i, are assumed to be independently beta (α, β)-distributed, and the outdegrees, a_i, are condition-

Figure 2: Simulated distributions of T_i and T_i', $i = 1, 2, 3, 4$, when p_i is beta $(0.2, 0.2)$ distributed and $n = 20$.

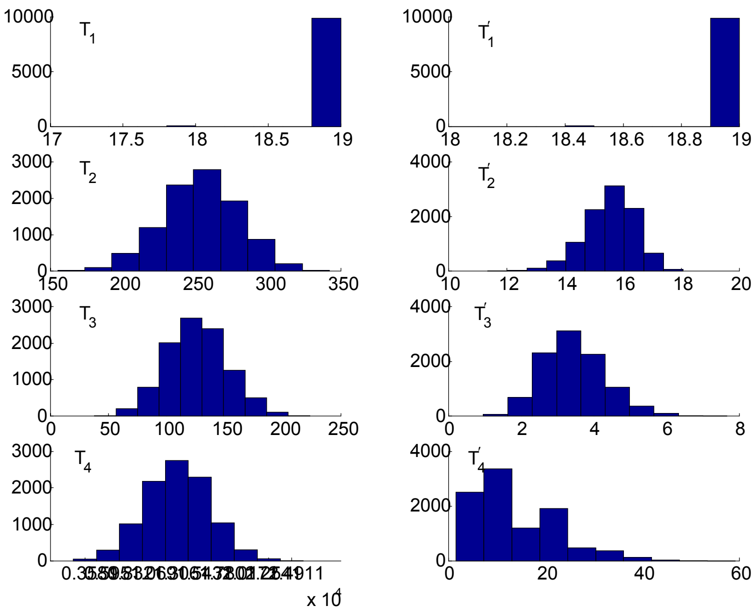

Figure 3: Simulated distributions of T_i and T_i', $i = 1, 2, 3, 4$, when p_i is beta $(1.0, 0.5)$ distributed and $n = 20$.

ally independent bin $(n-1, p_i)$-distributed. Then the expected value and variance of a_i is conveniently derived as

$$\begin{aligned} E(a_i) &= E[E(a_i|p_i)] = (n-1)E(p_i) = (n-1)\mu & (1) \\ Var(a_i) &= E[Var(a_i|p_i)] + Var[E(a_i|p_i)] & (2) \\ &= E[(n-1)p_i(1-p_i) + Var[(n-1)p_i]] \\ &= (n-1)\mu - (n-1)\left[\sigma^2+\mu^2\right] + (n-1)^2\sigma^2 \\ &= (n-1)\left[\mu(1-\mu) + (n-2)\sigma^2\right], \end{aligned}$$

where

$$\mu = \frac{\alpha}{\alpha+\beta} \tag{3}$$

and

$$\sigma^2 = Var(p_i) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}. \tag{4}$$

$E(a_i)$ is estimated by

$$\bar{a} = \frac{1}{n}\sum_i a_i, \tag{5}$$

and $Var(a_i)$ is estimated by

$$s_a^2 = \frac{1}{n}\sum_i (a_i - \bar{a})^2. \tag{6}$$

Note that the estimators given in (5) and in (6) are equal to T_2 and T_4 respectively. The moments of T_2 and T_4 are obtained from (1) and (2) in the following way:

$$E(T_2) = E(\bar{a}) = (n-1)\mu \tag{7}$$

$$Var(T_2) = Var(\bar{a}) = \frac{(n-1)}{n}\left[\mu(1-\mu) + (n-2)\sigma^2\right]$$

$$E(T_4) = E\left(s_a^2\right) = \frac{(n-1)}{n}Var(a_i) = \frac{(n-1)^2}{n}\left[\mu(1-\mu) + (n-2)\sigma^2\right] \tag{8}$$

If unbiased estimators of μ and σ^2 are needed it can be shown that (see Frank (2000))

$$\begin{aligned} E\left(\frac{\bar{a}}{n-1}\right) &= \mu \\ E\left\{\frac{n(n-1)-1}{(n-1)(n-2)}\frac{s_a^2}{(n-1)^2} - \frac{1}{n-2}\frac{\bar{a}}{n-1}\left(1-\frac{\bar{a}}{n-1}\right)\right\} &= \sigma^2 \end{aligned}$$

Guided by the results of the analytical derivations, the objective in this paper is to investigate relationships between T_2 and its corresponding closeness-based graph centrality measure T_2', and T_4 and its corresponding closeness-based graph centrality measure T_4'. Attempts are also made to find structures between T_1 and T_3 and their corresponding closeness-based graph centrality measures T_1' and T_3'.

4.2 Edge probabilities described by beta distribution parameters

To formalize the relationships between degree-based measures and closeness-based measures two approaches are taken. In this section the mean of each quantity T_i is compared to the mean of its corresponding quantity T_i', $i = 1, ..., 4$, as a function of the shape parameters in the beta-distribution, α and β. In Section 4.2 the comparisons are made when each quantity is regarded as a function of the mean and the variance of the probabilities for various beta-distributions.

Let $T_i(\alpha, \beta)$ and $T_i'(\alpha, \beta)$ be the degree-based quantity and the closeness-based quantity as a function of α and β. The values chosen for the simulations were $\alpha = 0.2, 0.5, 1.0, 1.5, 2.0, 4.0$ and $\beta = 0.2, 0.5, 1.0, 1.5, 2.0, 4.0$, which reflect different shapes of the beta distribution, of which two are illustrated in Figure 1.

10000 realizations of random directed graphs are simulated for each of the 36 combinations of the parameters α and β. In every graph, the eight quantities are evaluated. Then 36 averages of each quantity $T_i(\alpha, \beta)$ and $T_i'(\alpha, \beta)$ are calculated and denoted $m_i(\alpha, \beta)$ and $m_i'(\alpha, \beta)$ respectively. For notational simplicity, α and β is dropped and the average of the simulated measures as function of α and β is denoted m_i and m_i'. Values of m_i and m_i' are given in Appendix A.

Explicit functional relationships of the measures and the beta distribution parameters were derived for the mean of the graph centralities and the variance of the graph centralities for the degree-based actor centrality concept. To be able to study similar functional relationships for closeness-based

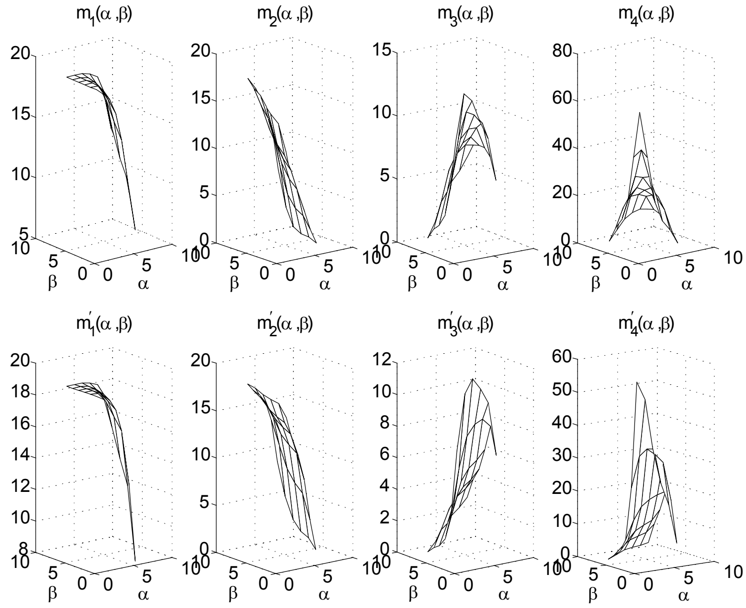

Figure 4: A plot of $m_i(\alpha, \beta)$ and $m'_i(\alpha, \beta)$, $i = 1, 2, 3, 4$ based on 10000 replications for each combination of α and β.

actor centrality measures no explicit formulas seem available. Closeness-based measures are therefore studied by simulations. The simulated results are provided in Figure 4 which shows the bivariate distributions of m_i and m'_i, $i = 1, ..., 4$. In general, it seems that there is a resemblance between the behaviour of any degree-based quantity, m_i, and the behaviour of the corresponding closeness-based quantity m'_i. The three-dimensional plots are difficult to interpret and therefore the quantities are plotted functions of α for various values of β in Figures 5-8.

For two of the degree-based measures, m_2 and m_4, explicit analytical expressions were derived in Section 4.1. According to Formula (7), the expected value of T_2 is given by

$$m_2 = (n-1)\,\mu = (n-1)\,\frac{1}{1+\frac{\beta}{\alpha}}, \tag{9}$$

implying that m_2 is increasing in α and decreasing in β. The relationships are confirmed in Figure 6a). Figure 6b) suggests that m'_2 could be approximated with m_2, at least for $\beta < 0.5$, and that also m'_2 is increasing in α and decreasing in β.

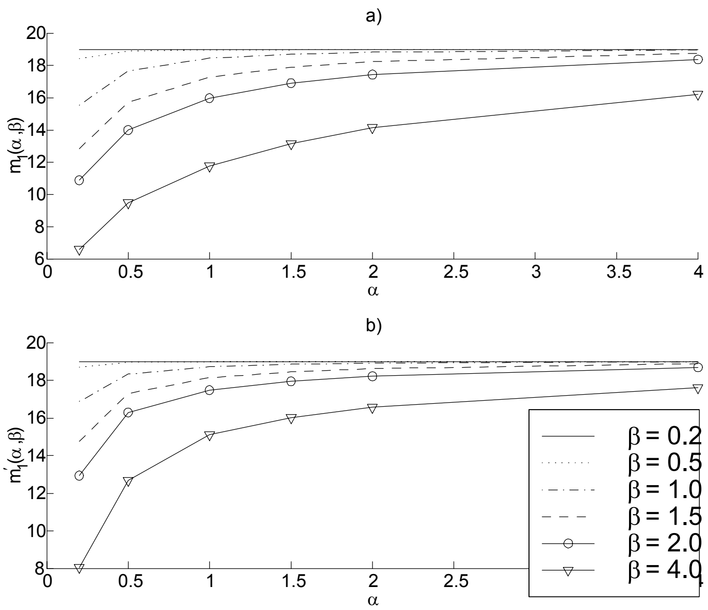

Figure 5: $m_1(\alpha, \beta)$ and $m_1'(\alpha, \beta)$ as functions of α for various β when $n = 20$.

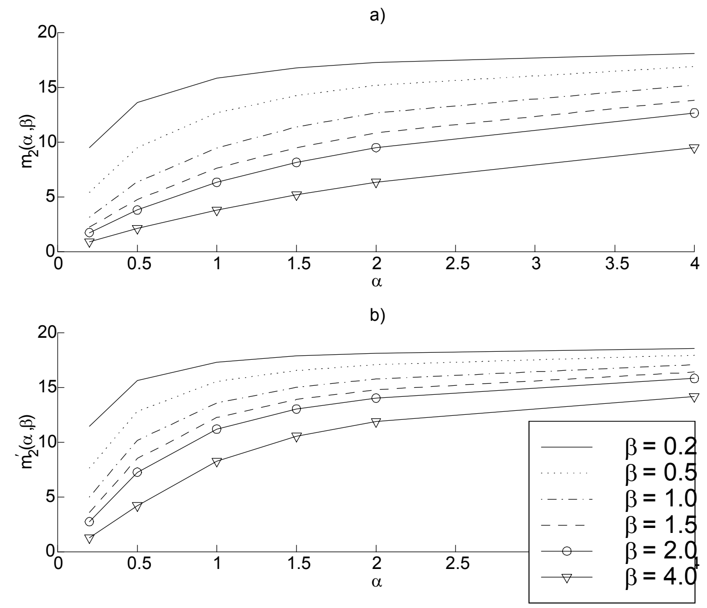

Figure 6: $m_2(\alpha, \beta)$ and $m_2'(\alpha, \beta)$ as functions of α for various β when $n = 20$.

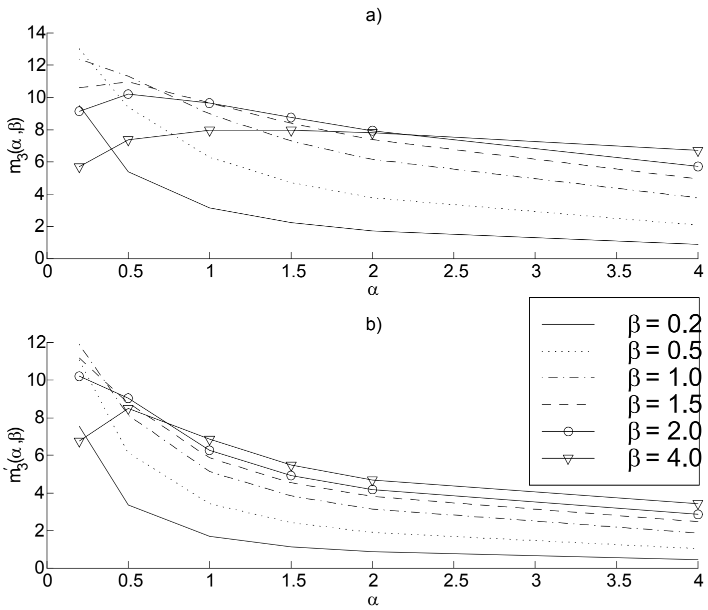

Figure 7: $m_3(\alpha,\beta)$ and $m'_3(\alpha,\beta)$ as functions of α for various β when $n = 20$.

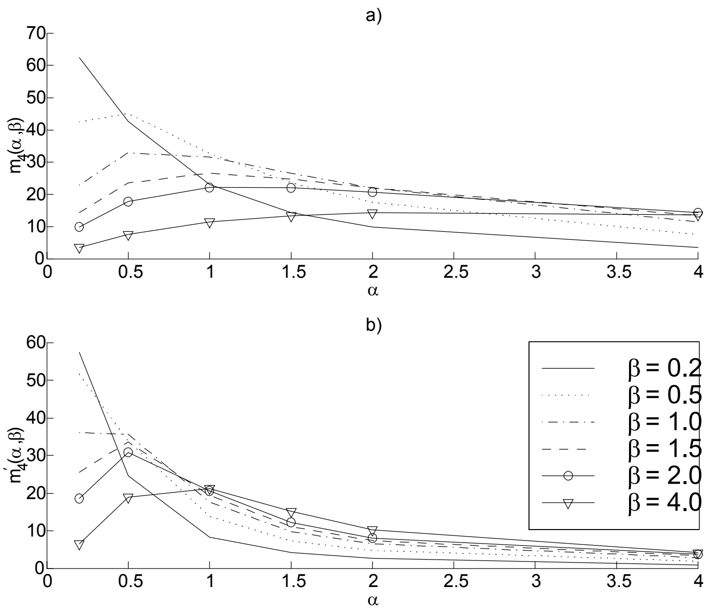

Figure 8: $m_4(\alpha,\beta)$ and $m'_4(\alpha,\beta)$ as functions of α for various β when $n = 20$.

β	α	$n^2\sigma^2$
0.2	0.16	72.6
0.5	0.34	52.4
1.0	0.62	36.1
1.5	0.88	27.6
2.0	1.14	22.3
4.0	2.15	12.7

Table 1: $m_4(\alpha, \beta)$ for various α and β.

From Formula (8) we have that the expected value of T_4 is given by

$$\begin{aligned} m_4 &= \frac{(n-1)^2}{n}\left[\mu(1-\mu) + (n-2)\sigma^2\right] \qquad (10) \\ &= \frac{(n-1)^2}{n}\left\{\frac{\alpha}{\alpha+\beta}\left(1 - \frac{\alpha}{\alpha+\beta}\right) + (n-2)\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}\right\}. \end{aligned}$$

Since m_4 is a more complex function of α and β, making a meaningful interpretation troublesome, a simplification is made. We see from Formula (10) that $n^2\sigma^2$ is its dominating part for graphs of sufficiently large order. Therefore we investigate the behaviour of m_4 by differentiating σ^2 as a function of α. The first differential is

$$\frac{d\sigma^2}{d\alpha} = -\beta\frac{2\alpha^2 + \alpha\beta + \alpha - \beta^2 - \beta}{(\alpha+\beta)^3(\alpha+\beta+1)^2} \qquad (11)$$

By setting equation (11) equal to zero and solving the equation, a maximum of σ^2 in the domain of α for fixed β, is obtained at

$$\alpha = -\frac{1}{4}\beta - \frac{1}{4} + \frac{1}{4}\sqrt{\left(9\beta^2 + 10\beta + 1\right)}, \quad \alpha > 0, \ \beta > 0. \qquad (12)$$

Maxima of $n^2\sigma^2$, for some fixed β and α - values obtained by (12), are given in Table 1. By comparing maxima of $n^2\sigma^2$ with the curves in Figure 8, it seems that $n = 20$ is sufficiently large to acceptably approximate m_4 with $n^2\sigma^2$.

For $\alpha = \beta$, the variance of the edge probabilities are given by $\sigma^2 = 1/4\,(2\beta+1)$, and for $\alpha \approx \beta$ an approximation of the variance is given by $\sigma^2 \approx 1/4\,(\alpha+\beta+1)$. This implies that for sufficiently large n, the variance of the out-degrees is given by

$$m_4 = \frac{n^2}{2\beta+1} \quad \text{for } \alpha = \beta$$

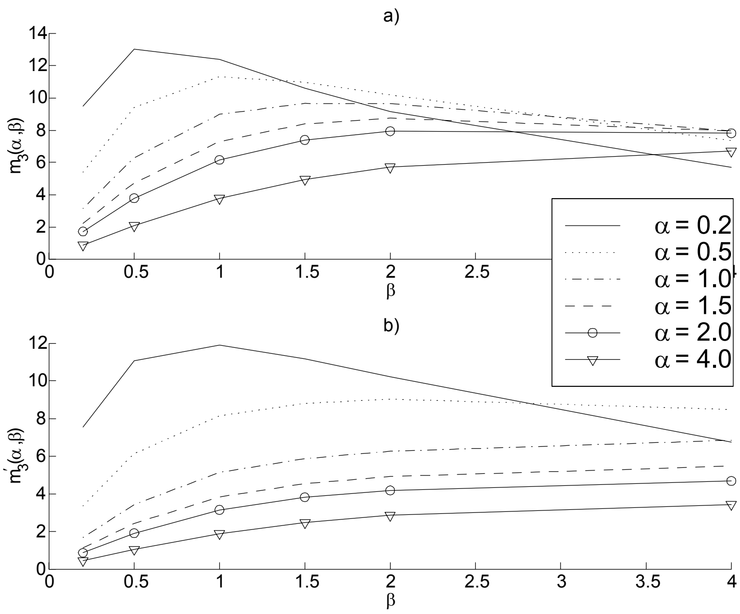

Figure 9: $m_3(\alpha, \beta)$ and $m_3'(\alpha, \beta)$ as functions of β for various α when $n = 20$.

or

$$m_4 = \frac{n^2}{\alpha + \beta + 1} \quad \text{for } \alpha \approx \beta.$$

Note that m_4/n^2 on the limit equals σ^2 as $n \longrightarrow \infty$. A sufficiently large number of replications have been made to achieve reliable results. By comparing the a) and b) parts of Figures 8 and 10, there seem to be no great differences between the variance centralities based on degree and on closeness. By investigating Figure 8, a slight difference is observed that might be crucial. m_4' decreases relatively rapidly in α, whereas m_4 has a smoother decreasing tendency.

We see in Figures 7 and 9, that the tendencies of m_3 and its corresponding closeness-based quantity, m_3', are difficult to interpret. An attempt is therefore made to investigate how the measures behave as functions of specific combinations of α and β. The measures are illustrated as functions of the ratio, β/α, in Figure 11. Furthermore, $m(T_3)$, $m(T_3')$, $m(T_4)$ and $m(T_4')$ are given as functions of the sum, $\alpha + \beta$, in Figure 12. The general trends of the degree-based measures and the corresponding closeness-based measures seem to agree. To draw extensive conclusions from the details appears to be risky.

According to Figure 5a), m_1 is an increasing function of α and it converges

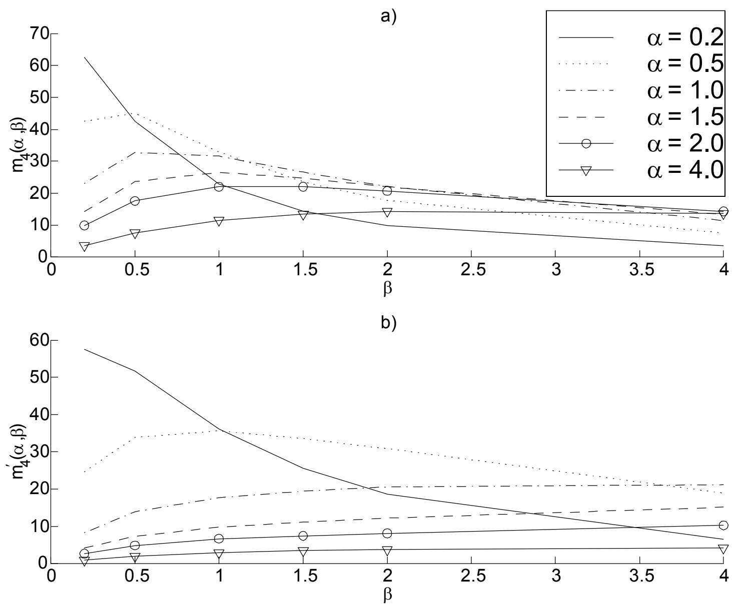

Figure 10: $m_4(\alpha, \beta)$ and $m_4'(\alpha, \beta)$ as functions of β for various α when $n = 20$.

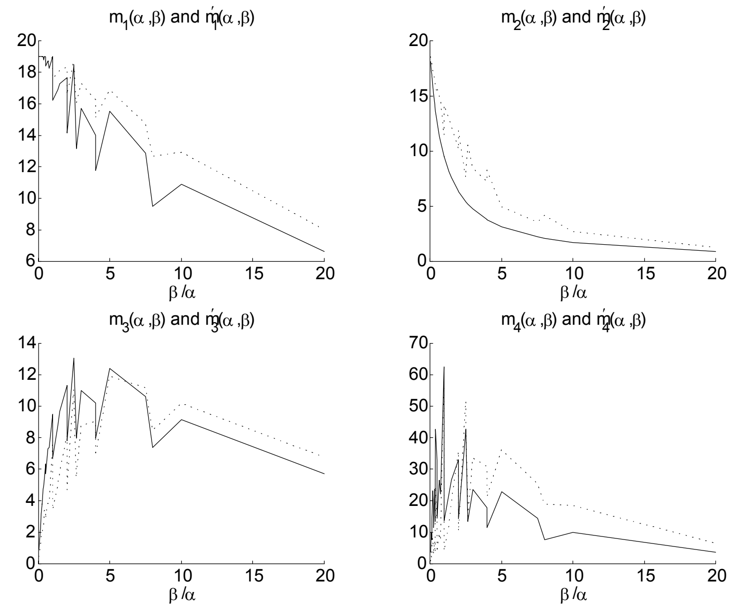

Figure 11: m_i (solid lines)and m_i' (dotted lines) plotted against β/α when $n = 20$.

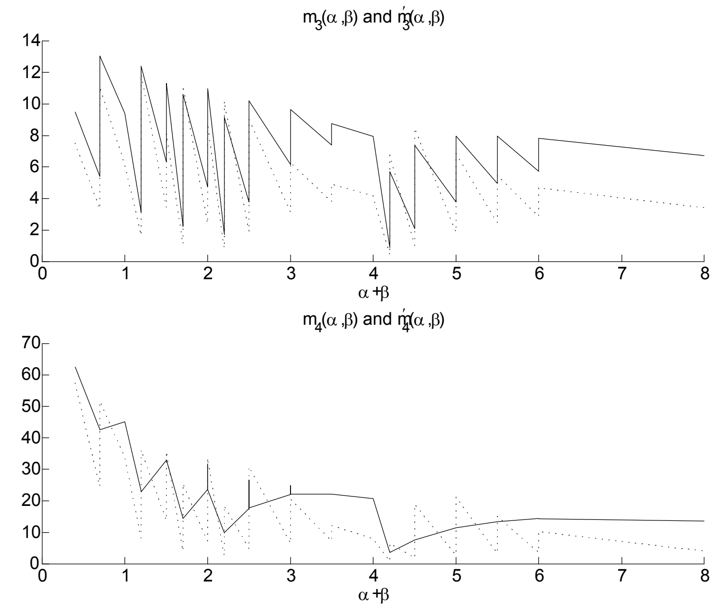

Figure 12: m_i (solid lines)and m_i' (dotted lines) plotted against $\alpha + \beta$ when $n = 20$.

to its maximum value $n - 1$ for various β. Conditioning on α, m_1 increases as β decreases. For small β, $\beta < 0.5$, m_1 attains approximately a constant value, $n - 1$, for all α. A similar discussion holds for m_1' according to Figure 5b).

4.3 Edge probabilities described by moments

A different aspect of the relationships between degree-based and closeness-based measures is attained by interpreting the measures as functions of the mean and the variance of p. Therefore an alternative approach is to consider the mean of the measures as functions of μ and σ^2, instead of as a functions of α and β. Thus in this section, $m_i\left(\mu, \sigma^2\right)$ and $m_i'\left(\mu, \sigma^2\right)$, $i = 1, 2, 3, 4$, is the mean of degree-based and closeness-based measures. $m_i\left(\mu, \sigma^2\right)$ and $m_i'\left(\mu, \sigma^2\right)$ are obtained by computer simulations. 10000 realizations of random graphs are generated. In each graph T_i and T_i' are evaluated numerically, and the mean of each measure is computed. As in the previous section we denote the means by m_i and m_i'.

The shape parameters of the beta-distribution can be expressed as linear functions of the mean , μ, and the variance, σ^2.

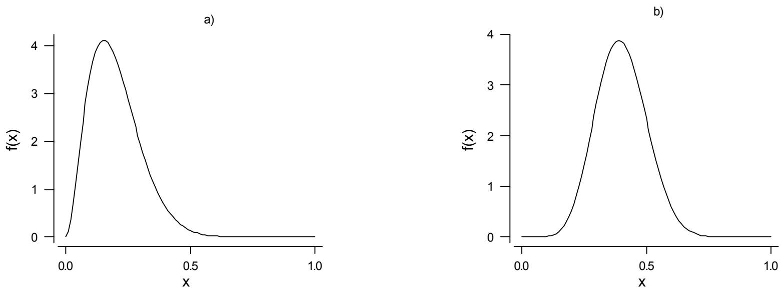

Figure 13: Densities of beta (α, β)-distributed variables with $\sigma^2 = 0.01$, $\mu = 0.2$ and $\mu = 0.4$; a) beta $(3, 12)$-distribution; b) beta $(9.2, 13.8)$-distribution.

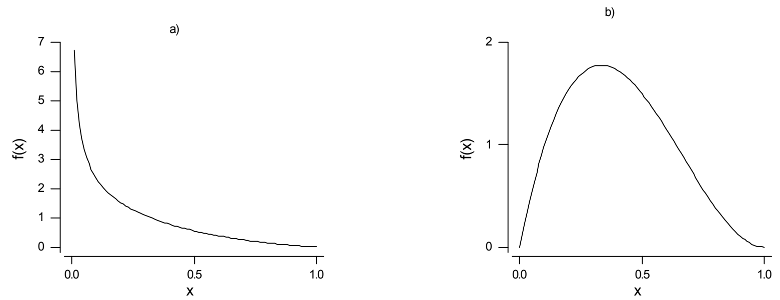

Figure 14: Densities of beta (α, β)-distributed variables with $\sigma^2 = 0.04$, $\mu = 0.2$ and $\mu = 0.4$; a) beta $(0.6, 2.4)$-distribution; b) beta $(2, 3)$-distribution.

α and β are related to μ and σ^2 through the following equations

$$\begin{aligned}
\alpha &= \frac{\mu^2 (1-\mu) - \mu\sigma^2}{\sigma^2} = \frac{\mu}{\sigma^2}\left[\mu(1-\mu) - \sigma^2\right] \\
\beta &= \frac{(1-\mu)^2 \mu - (1-\mu)\sigma^2}{\sigma^2} = \frac{(1-\mu)}{\sigma^2}\left[\mu(1-\mu) - \sigma^2\right],
\end{aligned}$$

obtained by rewriting (3) and (4). The distribution parameters α and β are determined for the eight combinations $\sigma^2 = 0.01, 0.04$ and $\mu = 0.2, 0.4, 0.6, 0.8$. Four of the densities of the beta distributions that have been used to simulate edge probabilities to obtain the desired σ^2 and μ, are shown in Figures 13 and 14.

Illustrations of the curves of m_i and m_i' are provided in Figures 15-18.

According to Figure 15, m_1 and m_1' are increasing for all μ and increasing for σ^2. Furthermore, m_1 and m_1' seem to be dependent on the variance for all μ, since there is a gap between the two curves (solid lines, $\sigma^2 = 0.01$; dotted lines, $\sigma^2 = 0.04$). A tendency of a decreasing gap of the curves as μ increases, indicates that the degree of the dependency of σ^2 is subsiding. As μ increases the curves converges to its maximum, $n-1$, independently of σ^2, which one intuitively would expect.

According to (7) in Section 4.1, m_2 is linear in μ and it is not depending on σ^2. This relationship agrees with the curves of m_2 provided in Figure 16, where m_2 exhibits a perfectly linear relationship in μ and the two curves for different σ^2 is coincident. Note that the two curves of m_2' are approximately coincident for $\mu > 0.4$, implying that, at least for $\mu > 0.4$, m_2' is linear in μ and coincident for various σ^2. For smaller values of μ there is a dependence on σ^2. Furthermore, both m_2 and m_2' are increasing in μ for all σ^2.

The measures based on heterogeneity show, in contrast to the other two measures, a non-increasing trend in μ. Both m_3 and m_3' are decreasing in μ according to Figure 17, which is intuitively correct. Since both T_1 and T_1' depend on σ^2, and due to the fact that T_3 and T_3' are combinations of the two measures, $T_1 - T_2$ and $T_1' - T_2'$ respectively, it follows that m_3 and m_3' are depending on σ^2. The tendencies are, analogously with T_1 and T_1', for T_3 and T_3' that the strength of the dependency decreases in μ.

According to Figure 18a), m_4 behaves as a second degree polynomial in μ, although the influence of μ is weak, that is symmetric about $\mu = 0.5$. Due to the fact that the influence of μ is weak implies that a good approximation could be done with a function constant of μ. Furthermore, the gap between the two curves in 18a) indicating that a factor containing σ^2 ought to be included, agrees with the mathematical relationship derived in Formula (8), see Section 4.1. In Section 4.2 it was pointed out that $m_4 \approx n^2\sigma^2$ for sufficiently large n, and $m_4 = (n-1)^2 \left[\mu(1-\mu) + (n-2)\sigma^2\right]/n$ for small n. According to Figure 18, the approximation, $n^2\sigma^2$, seems satisfactory for $n = 20$. Note, that the behaviour of the curves of m_4 and m_4' as functions of μ is different. While m_4 is approximately constant in μ, m_4' is decreasing. An intuitive explanation is as follows:

Assume that an edge is added from a vertex to any other vertex. Then the actor centrality index based on degree is only increasing for the two vertices involved. If a sufficiently large number of edges are added from a minority of actors, then the heterogeneity of the actor indices will increase leading to a large value of T_4. If we instead consider measuring centrality with closeness, then adding an edge from a vertex to any other vertex would not only increase the actor centrality index for the two vertices involved, but possibly for several other actors too. There is an indirect effect on the closeness-

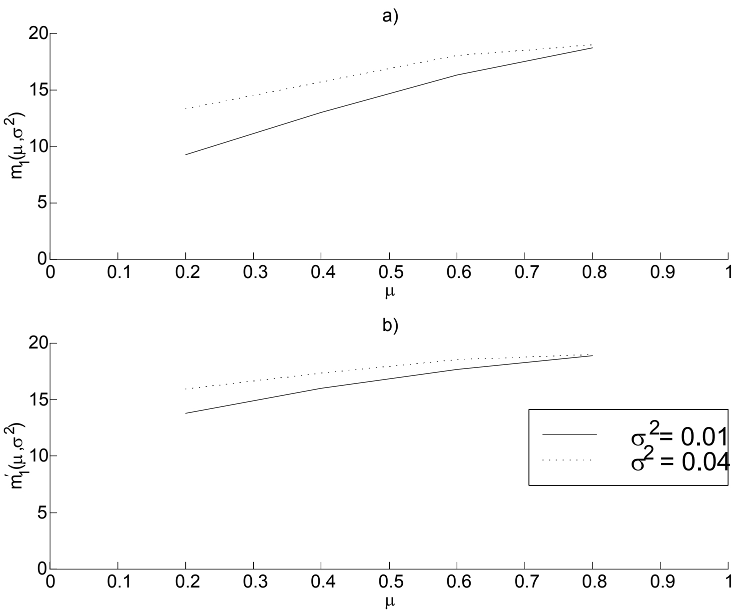

Figure 15: $m_1(\mu, \sigma^2)$ and $m_1'(\mu, \sigma^2)$ as functions of μ for various σ^2 when $n = 20$.

based actor centralities leading to that the distance between many nodpairs attain values one or two. Therefore the variance of c_i rapidly decreases when edges are added. A similar phenomenon is discussed in Tallberg (2000) where the power functions of T_4 and T_4' is compared when centrality is tested in non-directed random graphs.

5 Associations of centrality measures

A simple illustrative example of the associations between degree-based measures and their corresponding closeness-based measures are given in Figure 19. 1000 random graphs were generated according to the same distribution. The shape parameters of the beta distribution were chosen so that the difference between the degree-based measures and the closeness-based measures are sufficiently large, which occurs for small values of μ according to for example Figures 15-18. The mean of the edge probabilities was set to $\mu = 0.2$ and the variance of the edge probabilities was set to $\sigma^2 = 0.01$, i.e. p_i are independently beta$(3, 12)$-distributed. The general conclusion is the same as in the previous section. The associations T_1 versus T_1' and T_2 versus T_2' are relatively strong, while the associations between the degree-based

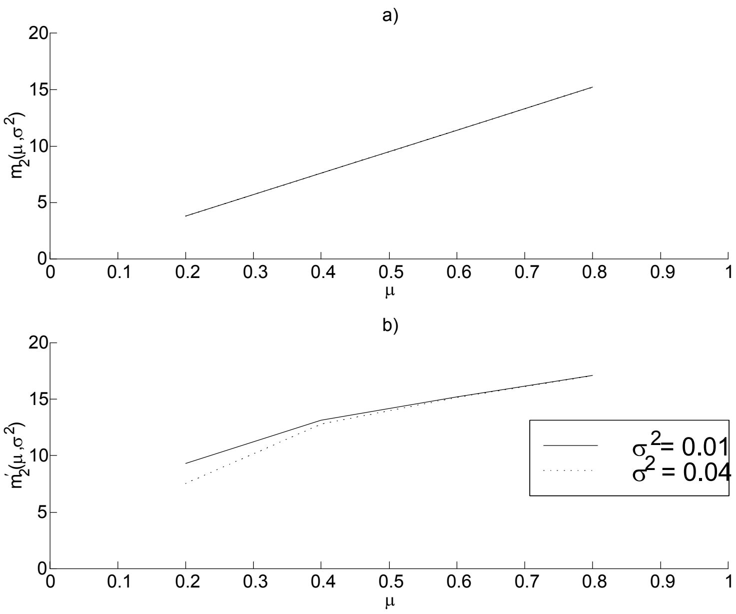

Figure 16: $m_2(\mu, \sigma^2)$ and $m_2'(\mu, \sigma^2)$ as functions of μ for various σ^2 when $n = 20$.

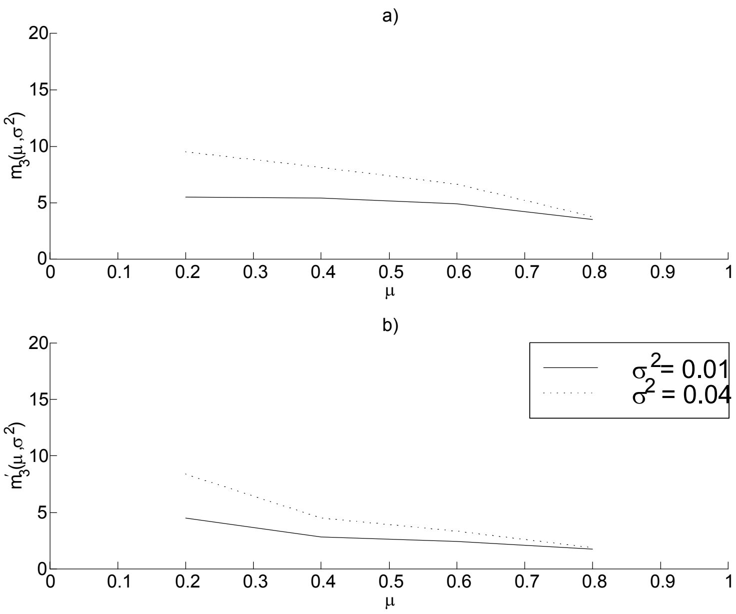

Figure 17: $m_3(\mu, \sigma^2)$ and $m_3'(\mu, \sigma^2)$ as functions of μ for various σ^2 when $n = 20$.

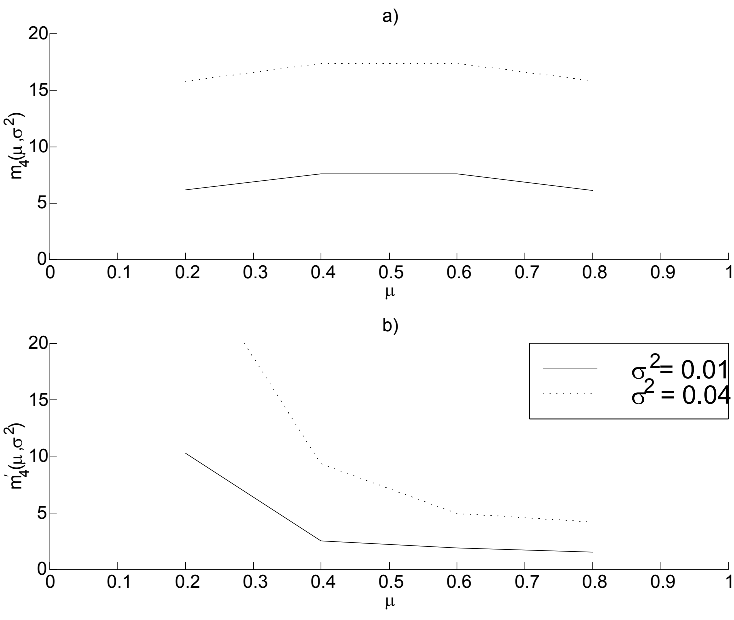

Figure 18: $m_4(\mu, \sigma^2)$ and $m_4'(\mu, \sigma^2)$ as functions of μ for various σ^2 when $n = 20$.

and closeness-based dispersion measures tend to be weaker. This is true in particular for the relationship between the variance of the actor centralities, T_4 versus T_4'. Table 2 provides descriptive statistics for the eight measures. Note that T_4 yields biased estimates of the variance of the outdegrees. The variance of the outdegrees, evaluated from Formula (2) given in Section 4.1, is $Var(a_i) = 6.46$, while the simulated mean of the estimator, T_4, yielded the value $m_4 = 6.137$. The result agrees with the derived expression in Formula (8), see Section 4.1. T_4 underestimates the variance of the outdegrees with a factor $(n-1)/n$.

The correlation coefficients of the degree-based measures and the corresponding closeness-based measures given in Table 2 reflects what is already stated by viewing the scatterplots in Figure 19. The correlations (T_i, T_i'), $i = 1, 2, 3$ are strong, while the correlation between the variance measures is weak. This result emphasizes what has already been mentioned in Section 4.2. That is, by investigating social networks within centrality contexts, using the maximum centrality, the mean centrality or the difference between the mean and the maximum centrality, we would probably capture similar features of the structures irrespective of centrality concept. By using the variance of the centralities, there is a significant risk that the analysis will yield crucially different result depending on the choice of concept, here

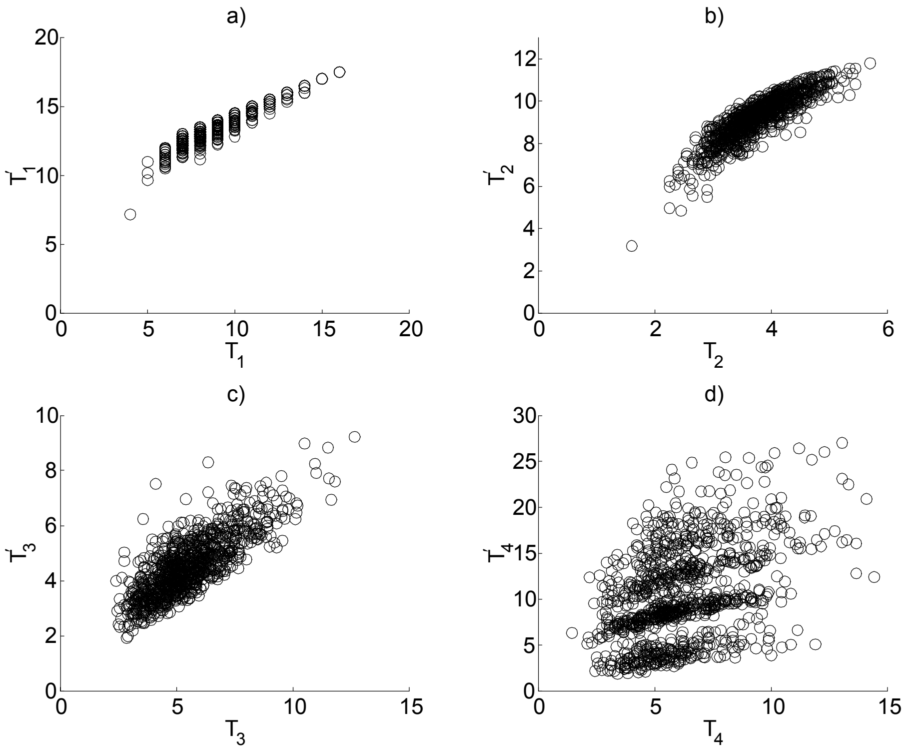

Figure 19: Degree-based measures versus closeness-based measures obtained by 1000 replications of random graphs of order $n = 20$ and where the edge probabilities are independent beta$(3, 12)$-distributed.

degree-based or closeness-based.

	Min	Max	Mean	St.Dev.	Corr(T_i, T_i')
T_1	4.0	16.0	9.3	1.8	0.95
T_1'	7.2	17.5	13.8	1.2	
T_2	1.6	5.7	3.8	0.6	0.88
T_2'	3.12	11.8	9.2	1.1	
T_3	2.4	12.7	5.5	1.6	0.80
T_3'	1.9	9.2	4.5	1.1	
T_4	1.4	14.4	6.1	2.1	0.44
T_4'	1.9	27.0	10.5	5.1	

Table 2: Descriptive statistics for the eight measures T_i and T_i', i = 1,2,3,4.

6 Concluding remarks

Statistical properties are investigated for three common graph centrality measures used in the field of social networks, and one measure introduced in Tallberg (2000). The properties are discussed and compared for two actor centrality concepts, degree centrality and closeness centrality. The model generating the edges in a social network of fixed order n is assumed to be conditionally independent Bernoulli(p_i)-distributed, where the edge probabilities, p_i, are independently beta-distributed. The graph centrality measures considered are the maximum actor centrality, the mean and the difference between the maximum and the mean of the actor centralities, and finally the variance of the actor centralities.

Simulated distributions of the quantities are obtained. Two different approaches are applied to investigate relationships between the degree-based graph centrality measures and the corresponding closeness-based graph centrality measures. The graph centrality measures are first considered as functions of the shape parameters of the beta distribution. It is observed that for the first three measures, the degree-based and closeness-based actor centralities yield similar results. For the fourth graph centrality measure, the two actor centrality concepts seem to give different results. The graph centrality measures are also considered as functions of the mean and variance of the edge probabilities. Analogous results are obtained implying that similar structural features of the social network are captured by the first three graph centrality measures for both the actor centrality concepts. This doesn't seem to hold for the fourth graph centrality measure.

References

[1] Bavelas, A. (1948). A mathematical model for group structures. *Human Organization,* **7**, 16-30.

[2] Frank, O. (2000). Using centrality modeling in network surveys. Department of Statistics, Stockholm University.

[3] Freeman, L.C. (1977). A set of measures of centrality based on betweeness. *Sociometry,* **40**, 35-41.

[4] Freeman, L.C. (1979). Centrality in social networks: Conceptual clarification. *Social Networks*, **1**, 215-239.

[5] Hagberg, J. (2000). Centrality testing and the distribution of the degree variance in Bernoulli graphs. Department of Statistics, Stockholm University.

[6] Nieminen, J. On centrality in a graph. *Scandinavian Journal of Psychology*, **15**, 322-336.

[7] Sabidussi, G. (1966). The centrality index of a graph. *Psychometrika*, **31**, 581-603.

[8] Snijders, T.A.B. (1981a). The degree variance: An index of graph heterogeneity. *Social Networks*, **3**, 163-174.

[9] Snijders, T.A.B. (1981b). Maximum value and null moments of the degree variance. TW-report 229. Department of Mathematics, University of Groningen.

[10] Tallberg, C. (2000). Testing centrality in random graphs. Research Report 2000:5. Department of Statistics, Stockholm University.

[11] Wasserman, S. and Faust, K. (1994). *Social Network Analysis: methods and applications*. New York: Cambridge University Press.

A Tables

The results from the simulation study in Section 4.1 are given here. The results are based on 10000 replications.

β	0.2	0.5	1.0	1.5	2.0	4.0
α						
0.2	9.50	5.43	3.17	2.24	1.73	0.90
0.5	13.57	9.50	6.33	4.75	3.80	2.11
1.0	15.83	12.67	9.50	7.60	6.33	3.80
1.5	16.76	14.25	11.40	9.50	8.14	5.18
2.0	17.27	15.20	12.67	10.86	9.50	6.33
4.0	18.10	16.89	15.20	13.82	12.67	9.50

Table 3: $E(a_i)$ for various α and β and n = 20.

α \ β	0.2	0.5	1.0	1.5	2.0	4.0
0.2	65.82	44.93	24.23	15.12	10.40	3.84
0.5	44.93	47.50	34.62	24.94	18.67	8.02
1.0	24.23	34.62	33.25	28.01	23.22	12.16
1.5	15.12	24.94	28.01	26.13	23.27	14.20
2.0	10.40	18.67	23.22	23.27	21.85	15.08
4.0	3.84	8.02	12.16	14.20	15.08	14.25

Table 4: $V\left(a_i\right)$ for various α and β and n = 20.

α \ β	0.2	0.5	1.0	1.5	2.0	4.0
0.2	19.00	18.45	15.55	12.85	10.88	6.61
0.5	19.00	18.92	17.67	15.73	14.01	9.50
1.0	19.00	18.99	18.47	17.29	15.98	11.77
1.5	19.00	19.00	18.72	17.91	16.91	13.15
2.0	19.00	19.00	18.84	18.26	17.44	14.16
4.0	19.00	19.00	18.98	18.78	18.38	16.22

Table 5: $m_1\left(\alpha,\beta\right)$ for various α and β and n = 20.

α \ β	0.2	0.5	1.0	1.5	2.0	4.0
0.2	9.51	5.41	3.16	2.24	1.73	0.91
0.5	13.60	9.51	6.35	4.75	3.81	2.12
1.0	15.83	12.68	9.47	7.61	6.32	3.79
1.5	16.76	14.26	11.40	9.51	8.15	5.18
2.0	17.27	15.20	12.67	10.85	9.49	6.34
4.0	18.10	16.89	15.20	13.82	12.66	9.50

Table 6: $m_2\left(\alpha,\beta\right)$ for various α and β and n = 20.

α \ β	0.2	0.5	1.0	1.5	2.0	4.0
0.2	9.49	13.04	12.39	10.62	9.15	5.71
0.5	5.40	9.42	11.32	10.98	10.20	7.38
1.0	3.17	6.31	9.00	9.68	9.65	7.98
1.5	2.24	4.74	7.32	8.40	8.76	7.97
2.0	1.73	3.80	6.17	7.40	7.94	7.82
4.0	0.90	2.11	3.78	4.95	5.72	6.72

Table 7: $m_3(\alpha, \beta)$ for various α and β and n = 20.

α \ β	0.2	0.5	1.0	1.5	2.0	4.0
0.2	62.52	42.54	22.92	14.43	9.97	3.65
0.5	42.62	45.14	32.99	23.56	17.78	7.67
1.0	23.10	32.89	31.69	26.62	22.15	11.49
1.5	14.39	23.65	26.54	24.84	22.09	13.45
2.0	9.95	17.63	22.02	22.08	20.73	14.35
4.0	3.64	7.62	11.54	13.46	14.39	13.65

Table 8: $m_4(\alpha, \beta)$ for various α and β and n = 20.

α \ β	0.2	0.5	1.0	1.5	2.0	4.0
0.2	19.00	18.71	16.89	14.76	12.94	8.05
0.5	19.00	18.96	18.33	17.31	16.30	12.68
1.0	19.00	18.99	18.73	18.14	17.48	15.12
1.5	19.00	19.00	18.86	18.45	17.95	16.03
2.0	19.00	19.00	18.92	18.63	18.22	16.57
4.0	19.00	19.00	18.99	18.89	18.69	17.61

Table 9: $m'_1(\alpha, \beta)$ for various α and β and n = 20.

α \ β	0.2	0.5	1.0	1.5	2.0	4.0
0.2	11.46	7.64	4.98	3.59	2.74	1.28
0.5	15.63	12.80	10.18	8.50	7.25	4.20
1.0	17.31	15.54	13.59	12.27	11.21	8.26
1.5	17.86	16.56	15.01	13.90	13.03	10.55
2.0	18.13	17.08	15.77	14.80	14.03	11.88
4.0	18.55	17.95	17.10	16.41	15.82	14.18

Table 10: $m_2'(\alpha, \beta)$ for various α and β and n = 20.

α \ β	0.2	0.5	1.0	1.5	2.0	4.0
0.2	7.54	11.07	11.91	11.17	10.20	6.76
0.5	3.37	6.16	8.15	8.81	9.04	8.48
1.0	1.69	3.45	5.15	5.87	6.28	6.86
1.5	1.14	2.44	3.85	4.55	4.93	5.48
2.0	0.87	1.92	3.15	3.83	4.19	4.69
4.0	0.45	1.05	1.89	2.48	2.87	3.43

Table 11: $m_3'(\alpha, \beta)$ for various α and β and n = 20.

α \ β	0.2	0.5	1.0	1.5	2.0	4.0
0.2	57.42	51.64	36.19	25.54	18.58	6.59
0.5	24.65	33.91	35.65	33.64	30.87	18.97
1.0	8.31	13.94	17.75	19.47	20.63	21.26
1.5	4.20	7.38	9.88	11.21	12.27	15.14
2.0	2.65	4.87	6.69	7.47	8.12	10.28
4.0	0.91	1.92	2.91	3.44	3.74	4.18

Table 12: $m_4'(\alpha, \beta)$ for various α and β and n = 20.

The results from the simulation study in Section 4.2 are given here. The results are based on 10000 replications.

		$m_1(\mu, \sigma^2)$		$m'_1(\mu, \sigma^2)$	
	σ^2	0.01	0.04	0.01	0.04
μ					
0.2		9.30	13.35	13.79	15.94
0.4		13.00	15.70	16.00	17.35
0.6		16.30	18.06	17.65	18.53
0.8		18.73	19.00	18.86	19.00

Table 13: $m_1(\mu, \sigma^2)$ and $m'_1(\mu, \sigma^2)$ for various μ and σ^2 and n = 20.

	$m_2(\mu,\sigma^2)$		$m'_2(\mu,\sigma^2)$	
σ^2	0.01	0.04	0.01	0.04
μ				
0.2	3.80	3.80	9.29	7.54
0.4	7.60	7.59	13.17	12.82
0.6	11.40	11.41	15.20	15.16
0.8	15.20	15.21	17.10	17.09

Table 14: $m_2(\mu,\sigma^2)$ and $m'_2(\mu,\sigma^2)$ for various μ and σ^2 and n = 20.

	$m_3(\mu,\sigma^2)$		$m'_3(\mu,\sigma^2)$	
σ^2	0.01	0.04	0.01	0.04
μ				
0.2	5.50	9.54	4.50	8.40
0.4	5.40	8.11	2.83	4.53
0.6	4.89	6.65	2.45	3.36
0.8	3.53	3.79	1.76	1.91

Table 15: $m_3(\mu,\sigma^2)$ and $m'_3(\mu,\sigma^2)$ for various μ and σ^2 and n = 20.

	$m_4(\mu,\sigma^2)$		$m'_4(\mu,\sigma^2)$	
σ^2	0.01	0.04	0.01	0.04
μ				
0.2	6.14	15.79	10.29	28.24
0.4	7.58	17.37	2.55	9.34
0.6	7.60	17.36	1.91	4.96
0.8	6.10	15.83	1.53	4.22

Table 16: $m_4(\mu,\sigma^2)$ and $m'_4(\mu,\sigma^2)$ for various μ and σ^2 and n = 20.