Essays on markets over random networks and learning in Continuous Double Auctions

van de Leur, M.C.W.

Chapter 2

Efficiency in Large Markets over Random Erdős-Rényi Networks

2.1 Introduction

Random graphs have been of interest since the seminal papers of Erdős and Rényi (1960, 1961). In these papers the random graph is introduced and phase transitions are derived as the number of vertices converges to infinity. The main result is that a phase transition occurs as the expected number of edges per vertex crosses the threshold value $\frac{1}{2}$. During such a phase transition the structure of the graph changes dramatically; up to the threshold the graph consists mainly of isolated trees whereas after the phase transition a giant component of positive measure arises. The work in the field of random graphs has been summarised in Alon and Spencer (2008), Bollobás (1982) and Janson et al. (2000).

The work of Erdős and Rényi on phase transitions in random graphs has not been thoroughly extended to bipartite graphs, which are graphs whose vertices can be divided in two disjoint sets in such a way that edges only occur between the sets. In this chapter we derive the phase transitions of a bipartite graph depending on the probability of an edge. We find a similar transition of the graph at the value 1; below the threshold the graph is a collection of mainly isolated spanning trees and after the transition a giant component emerges.

We consider a market over such a random bipartite graph, in which buyers and sellers are randomly linked with a certain probability. There is an equal number of buyers and sellers, who all desire to trade one unit of the good and we consider the case where the number of traders converges to infinity. For simplicity buyers assign a value of one to the good and sellers have a cost of zero. We assume that traders behave truthfully and bid one or ask zero. We study the maximal set of trades in the random bipartite graph, which depends on the characteristics of the different phases. For this so-called Maximum Matching problem many algorithms have been found, f.i. in Mucha and Sankowski (2004) and West (1999). We derive bounds on the expected efficiency, which under these simplifications can be calculated by dividing the expected number of trades in the maximum matching, by the number of traders on one side of the market. We derive an algorithm to construct all spanning trees and the distribution of the degree of the vertices. This allows for a development of tighter bounds on expected efficiency in the range consisting of mainly spanning trees.

The organisation of this chapter is as follows. The model and graph theory are considered in Section 2.2, followed by the phases of random bipartite graphs in Section 2.3. Section 2.4 derives bounds on expected efficiency in an infinitely large market over such networks. Finally, Section 2.5 concludes. The proofs are given in an appendix.

2.2 Model

We consider a market with n buyers and n sellers and we let n converge to infinity. Buyer i and seller j are linked with each other with a fixed probability p, independent of other links. Trade occurs only between linked traders. A related example of a market over networks is the spot exchange market studied in Gould et al. (2013a). In this market, traders provide a blocklist that excludes some traders on the opposite side of the market as possible trading partners. Trades are possible when both traders are not included in the blocklist of the other. The blocklist is used to protect against adverse selection and to control counterparty risk, and is thus considered exogenous. However, in contrast to the spot exchange market we assume that links are realised with equal probability and independently of each other.

Traders desire to obtain or sell one unit of a good. The valuation of a buyer equals one and the cost of a seller is zero, and there is complete information about valuations and costs. A buyer receives a profit equal to his valuation minus the transaction price after a trade, and zero otherwise. The profit of a seller that trades equals the transaction price minus his cost, otherwise the profit equals zero. We consider the maximal expected efficiency given the restrictions of the network structure and thus assume that traders are truthful and bid one or ask zero. Expected efficiency is defined as the maximal expected surplus under the network structure divided by the maximal surplus in a complete network. Because every trade results in the same surplus, it is sufficient to determine the fraction of transactions.

2.2.1 Graph theory

The market can be considered as a random bipartite graph, which is an extended Erdős-Rényi network. Two sets of labelled *vertices* V^1 and V^2 denote the sets of buyers and sellers and the set of *edges* E represents the links between traders. A graph is called *bipartite* when every edge connects a vertex in V^1 with a vertex in V^2. We consider the *number of edges* $N(n)$ as a function of the number of traders n on one side of the market; the *probability of an edge* equals $p = \mathbb{E}(\frac{N(n)}{n^2})$.

A graph G_2 is called a *subgraph* of G_1 if the vertices V_2^1 and V_2^2 of G_2 are subsets of the vertices V_1^1 and V_1^2 of G_1, and the edges E_2 of G_2 are a subset of the edges E_1 of G_1. A subgraph is called of *size* k,l if it is constructed from k and l labelled vertices. A subgraph is an *isolated* subgraph when either both or neither of the endpoints of an edge in E_1 belong to the subgraph, i.e. a vertex in the subgraph cannot be linked with a vertex outside the subgraph.

We define different types of subgraphs. A sequence of m attached edges is called a *path of size* m. A graph is *connected* if there is a path between every pair of vertices. A connected isolated subgraph of G_1 is denoted as a *component* of G_1. A *spanning tree* of size k,l occurs when k and

l vertices are connected by exactly $k+l-1$ edges. A *cycle* of size k,k occurs when k and k vertices are connected by at least $2k$ edges and a path of size $2k$ exists. In a *complete* bipartite graph an edge exists between any point in V^1 and any point in V^2.

Two graphs are *isomorphic* if there is a one-to-one correspondence between the vertices and the edges of both graphs. The *degree* of a graph G is the average number of edges of the vertices. A graph G is *balanced* if it contains no subgraph that has a larger degree than G itself.

As we consider asymptotic behaviour of the graph we often use the order of variables. The *little* o notation $a(n) = o\left(b(n)\right)$ denotes that $\lim_{n\to\infty} \frac{|a(n)|}{b(n)} = 0$ which indicates that $b(n)$ grows much faster than $a(n)$. Functions have the same growth rate when $\frac{|a(n)|}{b(n)}$ is bounded, which is indicated with the *big* O notation $a(n) = O\left(b(n)\right)$. Two functions are *similar*, denoted as $a(n) \sim b(n)$, when they are asymptotically equal and thus $\lim_{n\to\infty} \frac{a(n)}{b(n)} = 1$.

For a given property D^*, the function $D(n)$ is called a *threshold function* with respect to $N(n)$ if D^* almost surely (a.s.) is not satisfied when the ratio $\frac{N(n)}{D(n)}$ converges to zero, and a.s. is satisfied if the ratio converges to infinity:

$$\lim_{n\to\infty} \mathbb{P}_{n,N(n)}(D^*) = \left\{ \begin{array}{c|l} 0 & \text{if } \lim_{n\to\infty} \frac{N(n)}{D(n)} = 0 \\ 1 & \text{if } \lim_{n\to\infty} \frac{N(n)}{D(n)} = \infty. \end{array} \right.$$

2.3 Phase transitions bipartite graphs

We consider the phase transitions for a bipartite graph, similar to Erdős and Rényi (1960, 1961), as the probability of an edge increases. From phase to phase the network structure of the market changes abruptly. As the probability increases the market evolves from a collection of larger and larger spanning trees to a market that contains cycles; and eventually a giant central market emerges that contains a positive fraction of all traders. In the next section we derive tighter bounds when the market consists almost surely (a.s.) solely of spanning trees. We prove most theorems, shown in the appendix, by considering the number of edges N. The Law of Large Numbers implies that if p is of some order, the number of realised links N is of order pn^2 almost surely. Hence these results also hold for the generalised random bipartite graph.

Phase 1: $p = o(\frac{1}{n}) \Longleftrightarrow N = o(n)$

In this phase the random graph consists a.s. solely of connected subgraphs that are spanning trees (Th. 2.3.5). Hence, a.s. there are no cycles (Cor. 2.3.2). Spanning trees of size k,l only exist from the threshold $n^{2-\frac{k+l}{k+l-1}}$ on (Cor. 2.3.1). For $p \sim \rho n^{2-\frac{k+l}{k+l-1}}$ the number of spanning trees of size k,l follows a Poisson distribution with $\lambda = \frac{\rho^{k+l-1}k^{l-1}l^{k-1}}{k!l!}$ (Th. 2.3.2).

Hence during this phase the expected number of links per trader converges to zero and the market consists of infinitely many isolated submarkets up to a certain size.

Phase 2: $p \sim \frac{c}{n} \Longleftrightarrow N \sim cn,$ for $c \leq 1$

Besides spanning trees, also cycles occur in the graph. For $c < 1$, the probability that the bipartite graph contains at least one cycle equals $1 - \sqrt{1-c^2}\mathrm{e}^{\frac{c^2}{2}}$, which is strictly smaller than 1 (Th. 2.3.8). The number of cycles of size k,k follows a Poisson distribution with $\lambda = \frac{1}{2k}c^{2k}$ (Th. 2.3.3), whereas the number of isolated cycles of size k,k follows a Poisson distribution with $\lambda = \frac{1}{2k}(c\mathrm{e}^{c})^{2k}$ (Th. 2.3.4). The total expected number of cycles is given by $\frac{1}{2}\log(\frac{1}{1-c^2}) - \frac{c^2}{2}$ (Th. 2.3.7) and the expected number of vertices that belong to a cycle equals $\frac{c^4}{1-c^2}$ (Th. 2.3.9).

Even though cycles emerge, almost every vertex belongs to a spanning tree (Th. 2.3.6). Moreover, the total number of components is given by $n - N + O(1)$ (Th. 2.3.10) and hence almost every component is a spanning tree. The possible cycles in the bipartite graph are thus negligible. The maximum number of spanning trees of size k,l, $\frac{k^{l-1}l^{k-1}}{k!l!} \cdot (\frac{k+l-1}{k+l})^{k+l-1}\mathrm{e}^{-(k+l-1)}$, is attained for $p \sim \frac{1}{n} \cdot \frac{k+l-1}{k+l}$ (Th. 2.3.2). In this phase spanning trees of all sizes exist.

For $c = 1$ the graph almost surely contains a cycle (Th. 2.3.8) and the total number of cycles is of order $\frac{1}{2}\log(n)$ (Th. 2.3.7). The expected number of components is given by $n - N + O(\log(n))$ (Th. 2.3.10).

The expected number of links per trader in this phase is given by the value c. Almost every of $n - N + O(1)$ submarkets is a spanning tree and almost every trader is part of a spanning tree.

Phase 3: $p \sim \frac{c}{n} \Longleftrightarrow N \sim cn,$ for $c > 1$
As the expected number of edges exceeds 1 the structure of the bipartite graph undergoes a sudden change. The probability that a vertex belongs to a spanning tree is smaller than 1 and equals $\frac{x(c)}{c}$, where $x(c) = \sum_{v=1}^{\infty} \frac{v^{v-1}(ce^{-c})^v}{v!}$ and $v = k + l$ (Th. 2.3.6). The number of components is given by $\frac{2n}{c}\left(x(c) - \frac{x(c)^2}{2}\right)$ (Th. 2.3.10). The greatest component covers a set of vertices of positive measure, which follows directly from Blasiak and Durrett (2005).

The expected number of trading partners exceeds 1 and now a giant central market arises that covers a positive fraction of the total market. Around the central market smaller and smaller submarkets exist.

Phase 4: $pn \to \infty$
As the expected number of links converges to infinity, almost surely every trader is part of the central market. With probability zero small submarkets exist and thus the number of components is of order $O(1)$ (Th. 2.3.10).

2.4 Bounds on expected efficiency

The different phases determined in the previous section allow us to consider the restrictions of the network structure on the number of trades. Assuming truthful traders with equal valuations and costs, the number of trades corresponds directly to the extracted surplus. The expected maximal efficiency under the random network structure equals the expected maximum number of trades divided by n. The problem of calculating this expected maximal efficiency reduces to the Maximum Matching problem; this is a matching at which the number of trades is maximised. A matching in which all traders can trade is called a perfect matching. Many algorithms have been established to determine the maximum matching of a given bipartite graph.

Translating the network into a matrix allows for some necessary and sufficient conditions for a perfect matching. We can represent the network by a matrix, where rows correspond to the buyers and columns to the sellers. A value of 1 at place i,j denotes a link between buyer i and seller j; the value 0 denotes the absence of a link. A perfect matching is available iff either:

- All diagonal elements equal 1, possibly after permuting rows and/or columns.
- Every subset of sellers is linked to a subset of buyers with at least the same cardinality, often referred to as the Marriage theorem of Hall (1935).
- There does not exist a block of zeros of size $k \cdot l$ with $k + l > n$.

2.4.1 Example

The expected maximal efficiency is calculated exactly for $n = 1, ..., 4$ by determining for every number of existing links N the number of possibilities of having a certain number of maximal trades t. For example for $n = 2$ the $2^4 = 16$ possible realisations of the network are given in Fig. 2.1, where the two buyers are shown on top and the two sellers on the bottom.

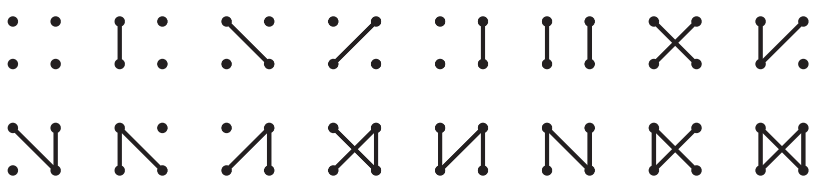


Figure 2.1: Possible network realisations for 2 buyers and 2 sellers.

We show the distribution of the maximal number of trades per number of links for $n = 3$ buyers and sellers; thus there are $2^9 = 512$ possible realisations of the network. For every possible realisation we determined the maximal number of trades and the number of links. The number of realisations with N links and a maximal possible number of trades t is shown in Table 2.1.

t \ N	0	1	2	3	4	5	6	7	8	9
0	1	0	0	0	0	0	0	0	0	0
1	0	9	18	6	0	0	0	0	0	0
2	0	0	18	72	90	45	6	0	0	0
3	0	0	0	6	36	81	78	36	9	1

Table 2.1: Example with 3 buyers and 3 sellers which shows the number of realisation of the network with N links and a maximal possible number of trades t.

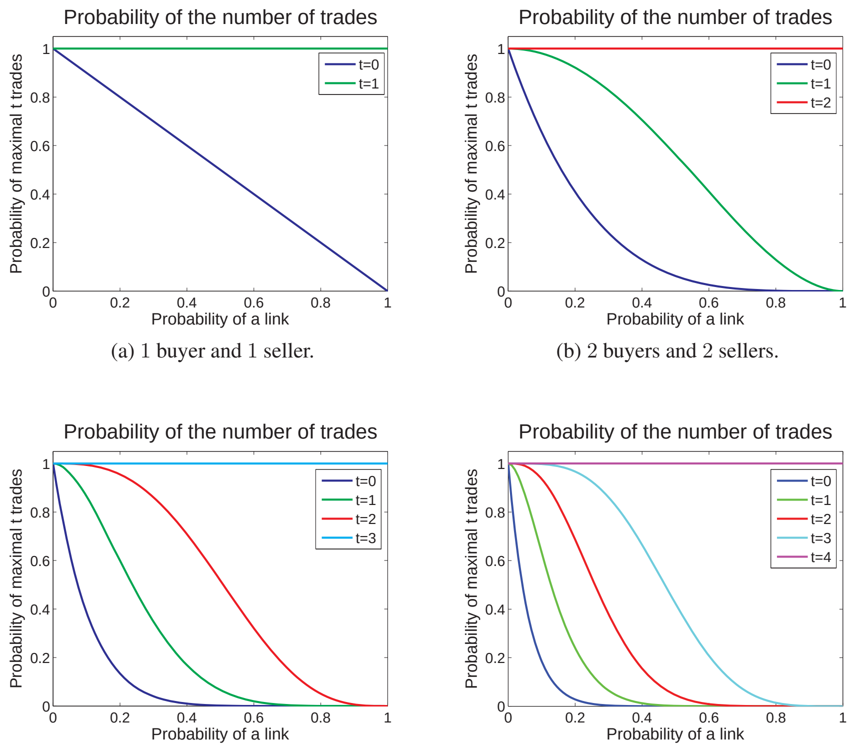
These calculations allow us to determine the distribution of the number of trades t for $n = 1, ..., 4$ buyers and n sellers, shown in Fig. 2.2.

For $n = 1, ..., 4$ buyers and sellers we show respectively the probability of full efficiency and the expected efficiency in Fig. 2.3. We observe that the probability of full efficiency increases for large p and decreases for small p. The expected efficiency is increasing in n because the expected number of links per trader increases.

2.4.2 Infinitely many traders

The expected maximal efficiency due to restrictions of the network structure is of interest in this section in a market with infinitely many traders. As mentioned before, this market is related to the spot exchange market studied in Gould et al. (2013a). For the different phases of the random bipartite graph expected efficiency for the entire market is calculated. As the expected number of links converges to zero, we find that the expected efficiency converges to zero. When the expected number of links however converges to a constant c we derive a lower bound of $1 - \frac{1-e^{-c}}{c}$ and an upper bound of $1 - e^{-c}$. Finally as the market becomes almost surely connected the probability that full efficiency is attained converges to one (Th. 2.4.1).

In the range $p = \frac{c}{n}$, $c \leq 1$ cycles are negligible and almost every vertex of the bipartite graph belongs to a spanning tree. Hence bounds for expected efficiency can be derived for spanning trees individually and added up. We formally show that the expected efficiency is continuous



(a) 1 buyer and 1 seller.

(b) 2 buyers and 2 sellers.

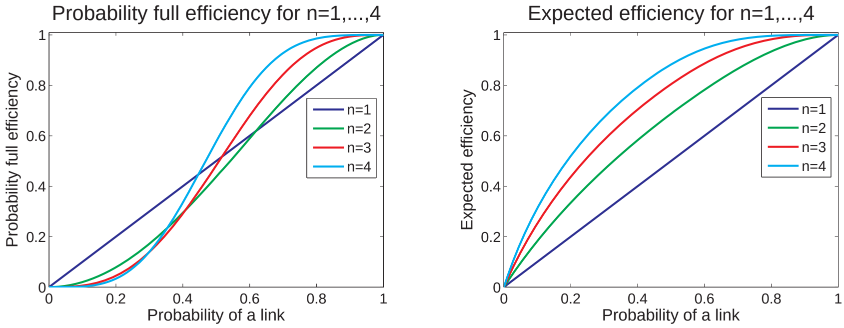
(c) 3 buyers and 3 sellers.

(d) 4 buyers and 4 sellers.

Figure 2.2: Distribution of the number of trades for $n = 1, ..., 4$ buyers and sellers. For every value of the probability of a link, the probability of maximal $t = 0, ..., n$ trades is given.

and especially at the point $c = 1$ of the phase transition (Th. 2.4.2).

In order to derive tighter bounds on expected efficiency in the range $p = \frac{c}{n}$, $c \leq 1$ we construct an algorithm that produces all possible, undirected, spanning trees of a certain size by adding vertices one by one to a directed tree. We show that this algorithm produces exactly all spanning trees (Th. 2.4.3).



(a) Probability of full efficiency. (b) Expected efficiency.

Figure 2.3: Efficiency as a function of the size of the market.

Algorithm 2.4.1: Constructing all possible spanning trees of a bipartite graph

All possible, undirected, spanning trees of size k,l can be constructed by forming a directed tree step by step. We denote the vertices by $V^1 = \{v_1^1, ..., v_k^1\}$ and $V^2 = \{v_1^2, ..., v_l^2\}$ respectively. The set of spanning trees is equivalent to the set of directed spanning trees with root v_1^1. This algorithm produces layer by layer all the possible spanning trees:

Step 1

The node v_1^1 is linked to a non-empty subset of V^2. This subset is removed from V^2 and v_1^1 is removed from V^1.

Step 2

All the vertices that are added to the directed tree in the last step are linked to a group of disjoint subsets of the other set of vertices that satisfy:

- The number of subsets in the group equals the number of vertices added in the last step.
- The union of the group of subsets is non-empty.
- The vertices linked to the same predecessor are ordered; to prevent counting isomorphic spanning trees multiple times.

The vertices in the group of subsets are removed from the set of vertices and this step is repeated until one set of remaining vertices is empty.

Step 3

All the vertices that are added to the directed tree in the last step are linked to a group of disjoint subsets of the other set of vertices that satisfy:

- The number of subsets in the group equals the number of vertices added in the last step.
- The union of the group is equal to the set of remaining vertices.
- The vertices linked to the same predecessor are ordered; to prevent counting isomorphic spanning trees multiple times.

This algorithm can easily be extended to multipartite graphs. In every step vertices are added that are a subset of the other sets of vertices. When all but one set of vertices is empty the algorithm moves on to Step 3. The distribution of the degrees of vertices in spanning trees can be determined using Algorithm 2.4.1. We show that every vertex in a spanning tree naturally has one edge and that the remaining edges are multinomially distributed per set of vertices (Th. 2.4.4).

This allows for tighter bounds on expected efficiency by considering the number of vertices with a degree larger than one, $\#V_{\text{degree}>1}$. This number of vertices can be calculated from the multinomial distribution of the remaining edges. We derive a lower bound of $\lceil \frac{\#V_{\text{degree}>1}+1}{2} \rceil$ on the expected efficiency in a spanning tree of size $k+l > 2$ and an upper bound of $min\,(k, l, \#V_{\text{degree}>1})$ (Th. 2.4.5). Together it can provide bounds on the expected maximal efficiency of the entire market when almost every component is a spanning tree. Considering spanning trees separately we find tighter bounds on expected maximal efficiency in the range $p \sim \frac{c}{n},\ c \leq 1$:
$\sum_{1\leq i\leq k\leq\infty} \sum_{1\leq j\leq l\leq\infty} \frac{\mathrm{e}^{-(k+l)}}{i!j!} S(l-1, k-i) S(k-1, l-j) \lceil \frac{k-i+l-j+1}{2} \rceil \leq \mathbb{E}(\text{eff})$
$\leq \sum_{1\leq i\leq k\leq\infty} \sum_{1\leq j\leq l\leq\infty} \frac{\mathrm{e}^{-(k+l)}}{i!j!} S(l-1, k-i) S(k-1, l-j) \min(k, l, k-i+l-j)$ (Th. 2.4.6).

These bounds are approximated by evaluating them for $k + l \leq 140$. Fig. 2.4 shows that these bounds are indeed tighter than the bounds found when the graph is considered as a whole.

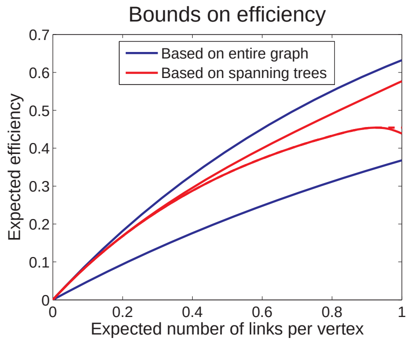


Figure 2.4: Bounds on expected efficiency on the basis of the entire graph and on the basis of spanning trees.

2.5 Concluding remarks

Following Erdős and Rényi (1960, 1961) we have constructed phase transitions for random bipartite graphs, where links are realised independently from each other with probability p. In the phase $p = o(\frac{1}{n})$ the graph consists of isolated spanning trees up to a certain size. The phase $p \sim \frac{c}{n}$, $c \leq 1$ is characterised by a graph where almost every component is a spanning tree. Such spanning trees occur of every size. The number of spanning trees follows a Poisson distribution. The greatest component is a spanning tree with zero measure. As c crosses the value 1 for $p \sim \frac{c}{n}$, the behaviour of the graph undergoes a sudden change. Besides spanning trees and small cycles, the graph contains a giant component of positive measure. As the expected number of edges per vertex converges to infinity, $p \cdot n \to \infty$, almost every vertex belongs to the giant component.

Using these phases we could find bounds for the expected efficiency in a market setting, for individual spanning trees and in general. We considered an equal number of buyers and sellers, who all desire to trade one unit of the good and consider the case where the number of traders converges to infinity. The results hold under the assumption that traders are truthful and bid or ask their valuation of 1 and cost 0 respectively. Under these settings the problem of finding the expected maximal efficiency reduces to the Maximum Matching problem. Moreover, the expected maximal efficiency can be calculated by dividing the expected number of trades in the maximum matching, by the number of traders on one side of the market. When the expected number of edges per vertex converges to zero or infinity, the expected efficiency converges to zero respectively one. In the range $p \sim \frac{c}{n}$ we have found a lower bound of $1 - \frac{1-\mathrm{e}^{-c}}{c}$ and an upper bound of $1 - \mathrm{e}^{-c}$ on expected efficiency.

These bounds can be improved in the range $p \sim \frac{c}{n},\ c \leq 1$ by considering the expected maximal efficiency of spanning trees separately. We introduced a new algorithm to construct all the spanning trees of a certain size and determined the distribution of the degrees of the vertices in spanning trees. In the phase where the bipartite graph consists mainly of spanning trees and other components can be neglected, the tighter bounds $\sum_{1\leq i\leq k\leq\infty}\sum_{1\leq j\leq l\leq\infty}\frac{\mathrm{e}^{-(k+l)}}{i!j!}S(l-1,k-i)S(k-1,l-j)\lceil\frac{k-i+l-j+1}{2}\rceil \leq \mathbb{E}(\text{eff})$
$\leq \sum_{1\leq i\leq k\leq\infty}\sum_{1\leq j\leq l\leq\infty}\frac{\mathrm{e}^{-(k+l)}}{i!j!}S(l-1,k-i)S(k-1,l-j)\min(k,l,k-i+l-j)$ are determined by considering the spanning trees separately.

As an extension of Erdős and Rényi (1960, 1961) we have found similar phase transitions for random bipartite graphs. Under an assumption about the distribution of links, random bipartite graphs describe the spot exchange market and we have derived bounds on expected maximal efficiency for every phase.

Appendix A: Theorems in Section 3

We prove most theorems by considering the number of edges N instead of the probability of a link p. The Law of Large Numbers implies that if p is of some order, the number of realised links N is of order pn^2 almost surely. Hence these results also hold for the generalised random bipartite graph. This random bipartite graph of size n,n with N edges is denoted as $\Gamma_{n,N}$.

Theorem 2.3.1

Let $k+l \geq 3$ and $k+l-1 \leq m \leq kl$ be positive integers. $\mathbb{B}_{k,l,m}$ denotes the non-empty set of connected balanced bipartite graphs of size k,l and m edges. The threshold function for the existence of at least one subgraph isomorphic with an element in $\mathbb{B}_{k,l,m}$ is $N = O(n^{2-\frac{k+l}{m}})$.

Proof

Let $\mathbf{B}_{k,l,m} \geq 1$ be the number of graphs in $\mathbb{B}_{k,l,m}$ that can be constructed from k and l labelled vertices. $\mathbb{P}_{n,N}(\mathbb{B}_{k,l,m})$ is the probability that the random graph $\Gamma_{n,N}$ contains at least one subgraph that is isomorphic to one of the elements in $\mathbb{B}_{k,l,m}$ and can be bounded by $\mathbb{P}_{n,N}(\mathbb{B}_{k,l,m})$ $\leq \binom{n}{k}\binom{n}{l}\mathbf{B}_{k,l,m}\frac{\binom{n^2-m}{N-m}}{\binom{n^2}{N}} = O(n^k n^l \frac{(n^2-m)^{N-m}}{(n^2)^N} \cdot \frac{N!}{(N-m)!}) = O(\frac{N^m}{n^{2m-k-l}})$. This holds since as $n \to \infty$, $\binom{n}{k} = \frac{n!}{(n-k)!k!} = O(n^k)$ for $k \geq 1$ and $\binom{n^2-m}{N-m} = \frac{(n^2-m)!}{(n^2-m-(N-m))!(N-m)!}$ $=O(\frac{(n^2-m)^{N-m}}{(N-m)!})$ for arbitrary N.

The k and l labelled vertices can be selected in $\binom{n}{k}\binom{n}{l}$ different ways and the m edges can form an element of $\mathbb{B}_{k,l,m}$ in $\mathbf{B}_{k,l,m}$ ways. The remaining $N-m$ edges can be selected from the remaining n^2-m possible edges. The above expression is only an upper bound, since graphs that contain multiple subgraphs isomorphic with an element of $\mathbb{B}_{k,l,m}$ are counted multiple times. Hence it remains to show that the graph contains a subgraph isomorphic with an element of $\mathbb{B}_{k,l,m}$ if N is at least of the order $n^{2-\frac{k+l}{m}}$.

We denote the set of all subgraphs S of $\Gamma_{n,N}$ that are isomorphic with an element of $\mathbb{B}_{k,l,m}$ by $\mathbb{B}^{(n)}_{k,l,m}$. Then $\mathbb{E}(\sum_{S\in\mathbb{B}^{(n)}_{k,l,m}} \mathbb{1}_{\{S\in\Gamma_{n,N}\}}) = \sum_{S\in\mathbb{B}^{(n)}_{k,l,m}} \mathbb{E}(\mathbb{1}_{\{S\in\Gamma_{n,N}\}}) = \binom{n}{k}\binom{n}{l}\mathbf{B}_{k,l,m}\frac{\binom{n^2-m}{N-m}}{\binom{n^2}{N}}$

$\sim \frac{\mathbf{B}_{k,l,m}}{k!l!} \cdot \frac{N^m}{n^{2m-k-l}}$.

For two elements $S_i, S_j \in \mathbb{B}^{(n)}_{k,l,m}$ that do not share an edge we find that $\mathbb{E}(\sum_{S_i,S_j\in\mathbb{B}^{(n)}_{k,l,m}} \mathbb{1}_{\{S_i,S_j\in\Gamma_{n,N}\}}) = \sum_{S_i,S_j\in\mathbb{B}^{(n)}_{k,l,m}} \mathbb{E}(\mathbb{1}_{\{S_i,S_j\in\Gamma_{n,N}\}}) \leq \binom{n}{2k}\binom{n}{2l}\mathbf{B}^2_{k,l,m}\frac{\binom{n^2-2m}{N-2m}}{\binom{n^2}{N}}$
$\leq \left(\binom{n}{k}\binom{n}{l}\mathbf{B}_{k,l,m}\frac{\binom{n^2-m}{N-m}}{\binom{n^2}{N}}\right)^2 \sim \left(\mathbb{E}(\sum_{S\in\mathbb{B}^{(n)}_{k,l,m}} \mathbb{1}_{\{S\in\Gamma_{n,N}\}})\right)^2$.

For two elements $S_i, S_j \in \mathbb{B}^{(n)}_{k,l,m}$ that share s,t vertices and $1 \leq r \leq m-1$ edges we find that $\mathbb{E}(\mathbb{1}_{\{S_i,S_j\in\Gamma_{n,N}\}}) = \frac{\binom{n^2-2m+r}{N-2m+r}}{\binom{n^2}{N}} = O(\frac{N^{2m-r}}{n^{4m-2r}})$.

Since all S_i are balanced the degree of the intersection of S_1 and S_2 should be less than the degree of the subgraph S_1 (and also S_2): $\frac{r}{s+t} \leq \frac{m}{k+l}$. Hence $s+t \geq \frac{r(k+l)}{m}$, and thus the number of such pairs of subgraphs S_i, S_j is bounded by $\mathbf{B}^2_{k,l,m}\sum_{s=1}^{k}\sum_{t=\frac{r(k+l)}{m}-s}^{l}\binom{n}{k}\binom{n}{l}\binom{k}{s}\binom{l}{t}\binom{n-k}{k-s}\binom{n-l}{l-t}$
$= O(\mathbf{B}^2_{k,l,m}\sum_{s=1}^{k}\sum_{t=\frac{r(k+l)}{m}-s}^{l}\frac{n^k n^l k^s l^t (n-k)^{k-s}(n-l)^{l-t}}{k!l!s!t!(k-s)!(l-t)!})$
$= O\left(\sum_{s=1}^{k}\sum_{t=\frac{r(k+l)}{m}-s}^{l} n^k n^l (n-k)^{k-s}(n-l)^{l-t}\right) = O(\sum_{s=1}^{k}\sum_{t=\frac{r(k+l)}{m}-s}^{l} n^{2(k+l)-s-t})$
$= O(n^{2(k+l)-\frac{r(k+l)}{m}})$, since $s+t \geq \frac{r(k+l)}{m}$. So $\mathbb{E}(\sum_{S_i,S_j\in\mathbb{B}^{(n)}_{k,l,m}} \mathbb{1}_{\{S_i,S_j\in\Gamma_{n,N}\}})$
$= O\left((\frac{N^m}{n^{2m-(k+l)}})^2\sum_{r=1}^{m-1}(\frac{n^{2-\frac{k+l}{m}}}{N})^r\right)$.

We combine the above results and find that
$\mathbb{E}\left((\sum_{S\in\mathbb{B}^{(n)}_{k,l,m}} \mathbb{1}_{\{S\in\Gamma_{n,N}\}})^2\right) = \mathbb{E}(\sum_{S_i,S_j\in\mathbb{B}^{(n)}_{k,l,m}} \mathbb{1}_{\{S_i,S_j\in\Gamma_{n,N}\}})$
$\leq \mathbb{E}(\sum_{S\in\mathbb{B}^{(n)}_{k,l,m}} \mathbb{1}_{\{S\in\Gamma_{n,N}\}}) + \left(\mathbb{E}(\sum_{S\in\mathbb{B}^{(n)}_{k,l,m}} \mathbb{1}_{\{S\in\Gamma_{n,N}\}})\right)^2 + O\left((\frac{N^m}{n^{2m-(k+l)}})^2\sum_{r=1}^{m-1}(\frac{n^{2-\frac{k+l}{m}}}{N})^r\right)$.

For $\frac{N^m}{n^{2m-k-l}} = \omega \to \infty$ it holds that
$\mathrm{Var}(\sum_{S\in\mathbb{B}^{(n)}_{k,l,m}} \mathbb{1}_{\{S\in\Gamma_{n,N}\}}) = \mathbb{E}\left((\sum_{S\in\mathbb{B}^{(n)}_{k,l,m}} \mathbb{1}_{\{S\in\Gamma_{n,N}\}})^2\right) - \left(\mathbb{E}(\sum_{S\in\mathbb{B}^{(n)}_{k,l,m}} \mathbb{1}_{\{S\in\Gamma_{n,N}\}})\right)^2$
$= \mathbb{E}(\sum_{S\in\mathbb{B}^{(n)}_{k,l,m}} \mathbb{1}_{\{S\in\Gamma_{n,N}\}}) + \left(\mathbb{E}(\sum_{S\in\mathbb{B}^{(n)}_{k,l,m}} \mathbb{1}_{\{S\in\Gamma_{n,N}\}})\right)^2 + O\left((\frac{N^m}{n^{2m-k-l}})^2\sum_{r=1}^{m-1}(\frac{n^{2-\frac{k+l}{m}}}{N})^r\right)$
$- \left(\mathbb{E}(\sum_{S\in\mathbb{B}^{(n)}_{k,l,m}} \mathbb{1}_{\{S\in\Gamma_{n,N}\}})\right)^2 = \frac{(\mathbb{E}(\sum_{S\in\mathbb{B}^{(n)}_{k,l,m}} \mathbb{1}_{\{S\in\Gamma_{n,N}\}}))^2}{\mathbb{E}(\sum_{S\in\mathbb{B}^{(n)}_{k,l,m}} \mathbb{1}_{\{S\in\Gamma_{n,N}\}})} + O\left((\frac{N^m}{n^{2m-k-l}})^2\sum_{r=1}^{m-1}(\frac{n^{2-\frac{k+l}{m}}}{N})^r\right)$
$= O\left(\frac{(\mathbb{E}(\sum_{S\in\mathbb{B}^{(n)}_{k,l,m}} \mathbb{1}_{\{S\in\Gamma_{n,N}\}}))^2}{\frac{N^m}{n^{2m-k-l}}}\right) = O\left(\frac{(\mathbb{E}(\sum_{S\in\mathbb{B}^{(n)}_{k,l,m}} \mathbb{1}_{\{S\in\Gamma_{n,N}\}}))^2}{\omega}\right)$.

Now we are able to use Chebysheff's inequality, which states that $\mathbb{P}(|X-\mu| \geq h\sigma) \leq \frac{1}{h^2}, \forall h > 0$. For $h = \frac{1}{2}\sqrt{\omega}$ we find that
$\mathbb{P}\left(|\sum_{S \in \mathbb{B}^{(n)}_{k,l,m}} \mathbb{1}_{\{S \in \Gamma_{n,N}\}} - \mathbb{E}(\sum_{S \in \mathbb{B}^{(n)}_{k,l,m}} \mathbb{1}_{\{S \in \Gamma_{n,N}\}})| \geq \frac{1}{2}\mathbb{E}(\sum_{S \in \mathbb{B}^{(n)}_{k,l,m}} \mathbb{1}_{\{S \in \Gamma_{n,N}\}})\right) = O(\frac{1}{\omega})$
$\Rightarrow \mathbb{P}\left(\sum_{S \in \mathbb{B}^{(n)}_{k,l,m}} \mathbb{1}_{\{S \in \Gamma_{n,N}\}} \leq \frac{1}{2}\mathbb{E}(\sum_{S \in \mathbb{B}^{(n)}_{k,l,m}} \mathbb{1}_{\{S \in \Gamma_{n,N}\}})\right) = O(\frac{1}{\omega})$.

As $\omega \to \infty$ we have that $\mathbb{E}(\sum_{S \in \mathbb{B}^{(n)}_{k,l,m}} \mathbb{1}_{\{S \in \Gamma_{n,N}\}}) \to \infty$ and thus a.s. $\Gamma_{n,N}$ contains a subgraph isomorphic to an element in $\mathbb{B}_{k,l,m}$ and the number of these subgraphs a.s. converges to ∞ with order of magnitude ω^m. □

Corollary 2.3.1

The threshold function for the existence of a spanning tree of size k,l with $m = k + l - 1$ edges is $N = O(n^{\frac{k+l-2}{k+l-1}})$.

Corollary 2.3.2

The threshold function for the existence of a connected subgraph of size k,l with $m = k + l \geq 3$ edges is $N = O(n)$. This connected subgraph contains precisely one cycle.

Corollary 2.3.3

The threshold function for the existence of a cycle of length $2k$ over k,k vertices with $m = 2k$ edges is $N = O(n)$, $k \geq 2$.

Corollary 2.3.4

The threshold function for the existence of a complete subgraph of size k,l with $m = k \cdot l$ edges is $p = O(n^{2-\frac{k+l}{k \cdot l}})$.

Lemma 2.3.1 (Erdős and Rényi, 1960)

Let $\epsilon_{n1}, \epsilon_{n2}, ..., \epsilon_{nl}$ be sets of l random variables on some probability space; suppose that ϵ_{ni} takes on only the values 1 or 0. If $\lim_{n\to\infty} \sum_{1 \leq i_1 < i_2 < ... < i_r \leq l} \mathbb{E}(\epsilon_{ni_1}, \epsilon_{ni_2}, ..., \epsilon_{ni_r}) = \frac{\lambda^r}{r!}$ uniformly in r for $r = 1, 2, ...$, where $\lambda > 0$ and the summation is extended over all combinations $(i_1, i_2, ..., i_r)$ of order r of the integers $1, 2, ..., l$, then $\lim_{n\to\infty} \mathbb{P}(\sum_{i=1}^{l} \epsilon_{n_i} = j) = \frac{\lambda^j \mathrm{e}^{-\lambda}}{j!}$, $(j = 0, 1, ...)$. I.e. the distribution of the sum $\sum_{i=1}^{l} \epsilon_{n_i}$ tends for $n \to \infty$ to the Poisson-distribution with mean value λ.

Theorem 2.3.2

For $\tau_{k,l}$, the number of isolated spanning trees of size k,l in $\Gamma_{n,N}$, and $\lim_{n\to\infty} \frac{N(n)}{n^{-\frac{k+l}{k+l-1}}} = \rho > 0$ it holds that: $\lim_{n\to\infty} \mathbb{P}_{n,N}(\tau_{k,l} = j) = \frac{\lambda^j \mathrm{e}^{-\lambda}}{j!}$, $(j = 0, 1, ...)$, with $\lambda = \frac{\rho^{k+l-1} k^{l-1} l^{k-1}}{k!l!}$. Moreover, the maximum number of spanning trees of size k,l of $n \frac{k^{l-1} l^{k-1}}{k!l!} (\frac{k+l-1}{k+l})^{k+l-1} \mathrm{e}^{-(k+l-1)}$ is attained for $N \sim n^{\frac{k+l-1}{k+l}}$.

Proof

We denote the set of all spanning trees of size k,l that are subgraphs of $\Gamma_{n,N}$ by $T_{k,l}^{(n)}$. The variable $\epsilon(S)$ takes on the value 1 if it is an isolated subgraph and 0 otherwise. We prove the theorem by applying Lemma 2.3.1 to $\sum_{S \in T_{k,l}^{(n)}} \epsilon(S)$, which requires us to show that all its conditions are fulfilled.

We find $\mathbb{E}(\epsilon(S)) = \frac{\binom{(n-k)(n-l)}{N-k-l+1}}{\binom{n^2}{N}} \sim (\frac{N}{n^2})^{k+l-1} \mathrm{e}^{-(k+l)\frac{N}{n}}$ by induction. For $k = 0$ and $l = 0$, $\frac{\binom{(n-k)(n-l)}{N-k-l+1}}{\binom{n^2}{N}}$ equals $\frac{(n^2)!N!(n^2-N)!}{(N+1)!(n^2-N-1)!(n^2)!} = \frac{n^2-N}{N+1} \sim \frac{n^2}{N}$, and thus the equality holds for $k = 0$ and $l = 0$. This is the basis of the induction. Dividing $\frac{\binom{(n-k)(n-l)}{N-k-l+1}}{\binom{n^2}{N}}$ by its limit $(\frac{N}{n^2})^{k+l-1} \mathrm{e}^{-(k+l)\frac{N}{n}}$ gives $\frac{\frac{\binom{(n-k)(n-l)}{N-k-l+1}}{\binom{n^2}{N}}}{(\frac{N}{n^2})^{k+l-1} \mathrm{e}^{-(k+l)\frac{N}{n}}} = \frac{(n^2-(k+l)n+kl)!N!(n^2-N)!}{(N-k-l+1)!(n^2-(k+l)n+kl-N+k+l-1)!(n^2)!} (\frac{n^2}{N})^{k+l-1} \mathrm{e}^{(k+l)\frac{N}{n}}$. We disregard the negligible order term. Now we use this to construct the following step of induction, by dividing this term by the subsequent term with $k+1$ and l (k and $l+1$ works symmetrical):

$$\frac{\frac{(n^2-(k+l)n+kl)!N!(n^2-N)!}{(N-k-l+1)!(n^2-(k+l)n+kl-N+k+l-1)!(n^2)!} (\frac{n^2}{N})^{k+l-1} \mathrm{e}^{(k+l)\frac{N}{n}}}{\frac{(n^2-((k+1)+l)n+(k+1)l)!N!(n^2-N)!}{(N-(k+1)-l+1)!(n^2-((k+1)+l)n+(k+1)l-N+(k+1)+l-1)!(n^2)!} (\frac{n^2}{N})^{(k+1)+l-1} \mathrm{e}^{((k+1)+l)\frac{N}{n}}}$$

$$= \frac{\frac{(n^2-(k+l)n+kl)!}{(N-k-l+1)!(n^2-(k+l)n+kl-N+k+l-1)!} (\frac{n^2}{N})^{k+l-1} \mathrm{e}^{(k+l)\frac{N}{n}}}{\frac{(n^2-((k+1)+l)n+(k+1)l)!}{(N-(k+1)-l+1)!(n^2-((k+1)+l)n+(k+1)l-N+(k+1)+l-1)!} (\frac{n^2}{N})^{(k+1)+l-1} \mathrm{e}^{((k+1)+l)\frac{N}{n}}}$$

$$\sim \frac{(n^2-(k+l)n+kl)^{N-k-l+1}}{(n^2-(k+l+1)n+(k+1)l)^{N-k-l}} \frac{1}{n^2} \mathrm{e}^{-\frac{N}{n}} \sim n^2 \frac{1}{n^2} \mathrm{e}^{-\frac{N}{n}} \sim \mathrm{e}^{-\frac{N}{n}} \to 1.$$

And thus the limit is shown for $k+1$ and l which concludes the induction. Hence the equation is proved for $\lim_{n\to\infty} \frac{N(n)}{n^{-\frac{k+l}{k+l-1}}} = \rho > 0$.

Moreover, for disjoint $S_1, ...S_r \in T_{k,l}^{(n)},\ k,l,r \geq 1$, it holds that $\mathbb{E}\big(\epsilon(S_1), ..., \epsilon(S_r)\big) = \frac{\binom{(n-rk)(n-rl)}{N-r(k+l-1)}}{\binom{n^2}{N}} \sim (\frac{N}{n^2})^{r(k+l-1)} \mathrm{e}^{-(k+l)r\frac{N}{n}}$ when all S_i are disjoint and zero otherwise.

An extended version of Cayley's formula states that from k and l labelled points, $k^{l-1}l^{k-1}$ different spanning trees can be formed. Hence summing over all possible r-tuples of spanning trees in $T_{k,l}^{(n)}$ gives $\sum \mathbb{E}\big(\epsilon(S_1), ..., \epsilon(S_r)\big) \sim k^{l-1}l^{k-1}\frac{\binom{n}{k}^r\binom{n}{l}^r}{r!}(\frac{N}{n^2})^{r(k+l-1)}\mathrm{e}^{-(k+l)r\frac{N}{n}}$
$\sim (\frac{k^{l-1}l^{k-1}}{k!l!})^r \frac{n^{(k+l)r}}{r!}(\frac{N}{n^2})^{r(k+l-1)}\mathrm{e}^{-(k+l)r\frac{N}{n}}$.

For $\lim_{n\to\infty} \frac{N(n)}{n^{-\frac{k+l}{k+l-1}}} = \rho > 0$ we can conclude that $\lim_{n\to\infty} \sum \mathbb{E}\big(\epsilon(S_1), ..., \epsilon(S_r)\big) = \frac{\lambda^r}{r!}$, $r = 1, 2, ...$ with λ defined as before. Hence we showed that Lemma 2.3.1 can be applied to $\tau_{k,l} = \sum_{S\in T_{k,l}^{(n)}} \epsilon(S)$.

Rewriting the above formula gives $\mathbb{E}(\tau_{k,l}) = \frac{n^2}{N} \cdot \frac{(\frac{N}{n}\mathrm{e}^{-\frac{N}{n}})^{k+l}k^{l-1}l^{k-1}}{k!l!} = n \cdot m_{k,l}(\frac{N}{n})$ with $m_{k,l}(t) = \frac{k^{l-1}l^{k-1}t^{k+l-1}\mathrm{e}^{-(k+l)t}}{k!l!}$. For k,l fixed we solve $\frac{\partial}{\partial t}m_{k,l}(t)$
$= \frac{k^{l-1}l^{k-1}}{k!l!}\mathrm{e}^{-(k+l)t}t^{k+l-2}\,(k+l-1-(k+l)t) = 0$ and hence the maximum is attained at $t = \frac{k+l-1}{k+l}$, or $N \sim n\frac{k+l-1}{k+l}$. This maximum equals $n\frac{k^{l-1}l^{k-1}}{k!l!}(\frac{k+l-1}{k+l})^{k+l-1}\mathrm{e}^{-(k+l-1)}$. □

Theorem 2.3.3

Let $\gamma_{k,k}$ be the number of cycles of size k,k as a subgraph of $\Gamma_{n,N}$. For $N(n) \sim cn,\ c > 0$ we find that $\lim_{n\to\infty} \mathbb{P}(\gamma_{k,k} = j) = \frac{\lambda^j\mathrm{e}^{-\lambda}}{j!},\ \lambda = \frac{1}{2k}(\frac{N}{n})^{2k}$.

Proof

There are $\frac{1}{2}k!(k-1)!$ possible cycles of size k,k. Thus $\mathbb{E}(\gamma_{k,k}) = \binom{n}{k}\binom{n}{k}\frac{1}{2}k!(k-1)!\frac{\binom{n^2-2k}{N-2k}}{\binom{n^2}{N}}$
$\sim \frac{1}{2} \cdot \frac{n!n!k!(k-1)!}{k!k!(n-k)!(n-k)!} \cdot \frac{(n^2-2k)^{N-2k}}{(n^2)^N} \cdot \frac{N!}{(N-2k)!} \sim \frac{1}{2k}n^kn^k\frac{(N)^{2k}}{(n^2)^{2k}} \sim \frac{1}{2k} \cdot (\frac{N}{n})^{2k}$.

For $C_{k,k}^{(n)}$ all cycles of size k,k, let $\epsilon(S)$, $S \in C_{k,k}^{(n)}$, be equal to 1 if S is a subgraph of $\Gamma_{n,N}$ and 0 otherwise. Similar to above we find that $\mathbb{E}\left(\epsilon(S)\right) = \frac{\binom{n^2-2k}{N-2k}}{\binom{n^2}{N}} \sim \left(\frac{N}{n^2}\right)^{2k}$ and $\mathbb{E}\left(\epsilon(S_1),...,\epsilon(S_r)\right) = \frac{\binom{n^2-2kr}{N-2kr}}{\binom{n^2}{N}} \sim \frac{(n^2-2kr)^{N-2kr}}{(n^2)^N} \cdot \frac{N!}{(N-2kr)!} \sim \left(\frac{N}{n^2}\right)^{2kr}$. Since there are $\frac{1}{2}k!(k-1)!$ of these cycles, if we sum over all possible r-tuples of these cycles we have: $\sum \mathbb{E}\big(\epsilon(S_1),...,\epsilon(S_r)\big)$ $\sim \frac{\left(\binom{n}{k}\binom{n}{k}\frac{1}{2}k!(k-1)!\right)^r}{r!}\left(\frac{N}{n^2}\right)^{2kr} \sim \frac{\left(\frac{1}{2}\cdot\frac{n!n!k!(k-1)!}{k!k!(n-k)!(n-k)!}\right)^r}{r!}\cdot\left(\frac{N}{n^2}\right)^{2kr} \sim \frac{(\frac{1}{2k}n^k n^k)^r}{r!}\cdot\left(\frac{N}{n^2}\right)^{2kr} = \frac{\left(\frac{1}{2k}(\frac{N}{n})^{2k}\right)^r}{r!} = \frac{\lambda^r}{r!}$ with $\lambda = \frac{1}{2k}\left(\frac{N}{n}\right)^{2k}$. Since $\gamma_{k,k} = \sum \epsilon(S)$ we can apply Lemma 2.3.1 and hence for $N(n) \sim cn$ the number of cycles of size k,k follows a Poisson distribution with $\lambda = \frac{1}{2k}\left(\frac{N}{n}\right)^{2k}$. □

Theorem 2.3.4

Let $\gamma_{k,k}^*$ be the number of isolated cycles or size k,k as a subgraph of $\Gamma_{n,N}$. For $N(n) \sim cn$, $c > 0$ we find that $\lim_{n\to\infty} \mathbb{P}(\gamma_{k,k}^* = j) = \frac{\lambda^j e^{-\lambda}}{j!}$, $\lambda = \frac{1}{2k}(ce^c)^{2k}$.

Proof

There are $\frac{1}{2}k!(k-1)!$ possible cycles of size k,k. $\mathbb{E}(\gamma_{k,k}^*) = \frac{1}{2}k!(k-1)!\binom{n}{k}\binom{n}{k}\frac{\binom{(n-k)^2}{N-2k}}{\binom{n^2}{N}}$ $\sim \frac{1}{2}\cdot\frac{n!n!k!(k-1)!}{k!k!(n-k)!(n-k)!}\left(\frac{N}{n^2}\right)^{2k} e^{-2k\frac{N}{n}} \sim \frac{1}{2k}n^k n^k\left(\frac{N}{n}\right)^{2k} e^{-2k\frac{N}{n}} \sim \frac{1}{2k}\left(\frac{N}{n}e^{-\frac{N}{n}}\right)^{2k}$. We can apply Lemma 2.3.1 again and thus the number of isolated cycles of size k,k follows a Poisson distribution with $\lambda = \frac{1}{2k}(ce^c)^{2k}$ for $N(n) \sim cn$. □

Theorem 2.3.5

For $N = o(n)$ and $n \to \infty$ the graph $\Gamma_{n,N}$ is a.s. the union of disjoint spanning trees.

Proof

Let T be the property that a graph is the union of disjoint spanning trees and thus $\bar{T}$ the property that the graph contains at least one cycle. Then $\mathbb{P}_{n,N}(\bar{T}) \leq \sum_{k=2}^{n}\binom{n}{k}\binom{n}{k}\frac{1}{2}k!(k-1)!\frac{\binom{n^2-2k}{N-2k}}{\binom{n^2}{N}}$ $= O\left(\sum_{k=2}^{n}\frac{n!n!\frac{1}{2}k!(k-1)!}{k!k!(n-k)!(n-k)!}\cdot\frac{(n^2-2k)^{N-2k}}{(n^2)^N}\cdot\frac{N!}{(N-2k)!}\right) = O(\sum_{k=2}^{n}\frac{n^k n^k}{(n^2)^{2k}}N^{2k}) = O\left(\sum_{k=2}^{n}(\frac{N}{n})^{2k}\right)$ $= O(\frac{N}{n})$. Thus for $N = o(n)$ we have that $\lim_{n\to\infty}\mathbb{P}_{n,N}(T) = 1$. □

Theorem 2.3.6

Let $V_{n,N}$ be the number of vertices in $\Gamma_{n,N}$ which belong to an isolated spanning tree contained in $\Gamma_{n,N}$. For $N(n) \sim cn$ we have $\lim_{n\to\infty} \frac{\mathbb{E}(V_{n,N})}{2n} = 1$ when $c \leq 1$ and $\lim_{n\to\infty} \frac{\mathbb{E}(V_{n,N})}{2n} = \frac{x(c)}{c}$ when $c > 1$, where $x(c) = \sum_{v=1}^{\infty} \frac{v^{v-1}(ce^{-c})^v}{v!}$ and $v = k + l$.

Proof

For $\tau_{k,l}$ the number of isolated spanning trees of size k,l as a subgraph of $\Gamma_{k,l}$, then we have: $V_{n,N} = \sum_{k=0}^{n} \sum_{l=0}^{n} (k+l)\tau_{k,l}$ and hence $\mathbb{E}(V_{n,N}) = \sum_{k=0}^{n} \sum_{l=0}^{n} (k+l)\mathbb{E}(\tau_{k,l})$.

Using $\lim_{n\to\infty} \frac{\mathbb{E}(\tau_{k,l})}{2n} = \frac{1}{2c} \cdot \frac{(ce^{-c})^{k+l} k^{l-1} l^{k-1}}{k!l!}$ we obtain for $c \leq 1$ that $\lim_{n\to\infty} \frac{\mathbb{E}(V_{n,N})}{2n}$ $= \frac{1}{2c} \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \frac{(k+l)(ce^{-c})^{k+l} k^{l-1} l^{k-1}}{k!l!}$.

Remains to show that the latter term equals 1. We make use of $\sum_{k=0}^{v} \binom{v}{k} k^{v-k-1}(v-k)^{k-1}$ $= 2v^{v-2} \Rightarrow \sum_{k=0}^{v} \frac{k^{v-k-1}(v-k)^{k-1}}{k!(v-k)!} = \frac{2v^{v-2}}{v!}$ and $\sum_{v=1}^{\infty} \frac{v^{v-1}(ce^{-c})^v}{v!} = c$ for $c \leq 1$, see Erdős and Rényi (1960).

Thus $\lim_{n\to\infty} \frac{\mathbb{E}(V_{n,N})}{2n} = \frac{1}{2c} \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \frac{(k+l)(ce^{-c})^{k+l} k^{l-1} l^{k-1}}{k!l!}$
$= \frac{1}{2c} \sum_{v=1}^{\infty} \sum_{k+l=v} \frac{(k+l)(ce^{-c})^{k+l} k^{l-1} l^{k-1}}{k!l!} = \frac{1}{2c} \sum_{v=1}^{\infty} \sum_{k=0}^{v} \frac{v(ce^{-c})^v k^{v-k-1}(v-k)^{k-1}}{k!(v-k)!}$
$= \frac{1}{2c} \sum_{v=1}^{\infty} v(ce^{-c})^v \sum_{k=0}^{v} \frac{k^{v-k-1}(v-k)^{k-1}}{k!(v-k)!} = \frac{1}{2c} \sum_{v=1}^{\infty} v(ce^{-c})^v \frac{2v^{v-2}}{v!} = \frac{1}{c} \sum_{v=1}^{\infty} \frac{v^{v-1}(ce^{-c})^v}{v!}$
$= 1$. So a.e. vertex belongs to a spanning tree for $c \leq 1$.

Similarly to Erdős and Rényi (1960), for $c > 1$ it holds that $\lim_{n\to\infty} \frac{\mathbb{E}(V_{n,N})}{2n}$ $= \frac{1}{c} \sum_{v=1}^{\infty} \frac{v^{v-1}(ce^{-c})^v}{v!} < 1$. $\square$

Theorem 2.3.7

The total number of cycles in $\Gamma_{n,N}$ is denoted by $C_{n,N}$. For $N(n) \sim cn,\ c < 1$ it holds that $\lim_{n\to\infty} \mathbb{E}(C_{n,N}) = \frac{1}{2}\log(\frac{1}{1-c^2}) - \frac{c^2}{2}$ and for $c = 1$ that $\lim_{n\to\infty} \mathbb{E}(C_{n,N}) \sim \frac{1}{2}\log(n)$.

Proof

For $c < 1$ we find that $\mathbb{E}(\gamma_{k,k}) = \binom{n}{k}\binom{n}{k}\frac{1}{2}k!(k-1)!\frac{\binom{n^2-2k}{N-2k}}{\binom{n^2}{N}} \sim \frac{n!n!\frac{1}{2}k!(k-1)!}{k!k!(n-k)!(n-k)!} \cdot \frac{(n^2-2k)^{N-2k}}{(n^2)^N} \cdot \frac{N!}{(N-2k)!}$ $\sim \frac{1}{2k}(\frac{N}{n})^{2k} \sim \frac{1}{2k}c^{2k}$. Since $C_{n,N} = \sum_{k=2}^{n} \gamma_{k,k}$ and $\sum_{k=1}^{\infty} \frac{z^k}{k} = -\log(1-z)$ for $z = c^2$ we find that $\lim_{n\to\infty} \mathbb{E}(C_{n,N}) = \frac{1}{2}\log(\frac{1}{1-c^2}) - \frac{c^2}{2}$, which proves the first part.

For $c = 1$ it holds that $\mathbb{E}(\gamma_{k,k})$ is similar to $\frac{1}{2k}$ and hence $C_{n,N} = \sum_{k=2}^{n} \gamma_{k,k} = \frac{1}{2}\sum_{k=2}^{n} \frac{1}{k}$. Since $\sum_{k=1}^{n} \frac{1}{k} \sim \log(n+1)$ we can conclude that $\lim_{n\to\infty} \mathbb{E}(C_{n,N}) \sim \frac{1}{2}\log(n)$. □

Theorem 2.3.8

Let C be the property that a bipartite graph contains at least one cycle. When $N(n) \sim cn,\ c \leq 1$, it holds that $\lim_{n\to\infty} \mathbb{P}_{n,N}(C) = 1 - \sqrt{1-c^2}\mathrm{e}^{\frac{c^2}{2}}$. For $c < 1$ the probability of at least one cycle is less than 1, but for $c = 1$ the bipartite graph a.s. contains a cycle.

Proof

Given that the probability that two cycles are not disjoint is negligibly small and that the number of cycles follows a Poisson distribution with mean $\lambda = \lim_{n\to\infty} \mathbb{E}(C_{n,N})$, we find for $c < 1$ that $\lim_{n\to\infty} \mathbb{P}_{n,N}(\bar{C}) = \mathrm{e}^{-\lim_{n\to\infty} \mathbb{E}(C_{n,N})} = \mathrm{e}^{-\left(\frac{1}{2}\log(\frac{1}{1-c^2}) - \frac{c^2}{2}\right)} = \sqrt{1-c^2}\mathrm{e}^{\frac{c^2}{2}}$. As this converges to 0 as $c \uparrow 1$, the theorem is proved for $c \leq 1$. □

Theorem 2.3.9

The number of points of $\Gamma_{n,N}$ that are part of a cycle is denoted as $C^*_{n,N}$. For $N(n) \sim cn,\ c < 1$, it holds that $\lim_{n\to\infty} \mathbb{E}(C^*_{n,N}) = \frac{c^4}{2(1-c^2)}$.

Proof

Given that the probability that two cycles are not disjoint is negligibly small we find that $\lim_{n\to\infty} \mathbb{E}(C^*_{n,N}) \sim \lim_{n\to\infty} \sum_{k=2}^{n} 2k\gamma_{k,k} = c^4 \sum_{k=0}^{\infty}(c^2)^k = c^4\frac{1}{1-c^2} = \frac{c^4}{1-c^2}$, since $\sum_{k=0}^{n} z^k = \frac{1-z^{n+1}}{1-z}$. □

Theorem 2.3.10

We denote the number of components of $\Gamma_{n,N}$ by $\zeta_{n,N}$. For $N(n) \sim cn,\ c < 1$ it holds that $\mathbb{E}(\zeta_{n,N}) = n - N + O(1)$. When $c = 1$ the expected number of components is given by $\mathbb{E}(\zeta_{n,N}) = n - N + O\left(\log(n)\right)$. For $c > 1$ we find $\lim_{n\to\infty} \frac{\mathbb{E}(\zeta_{n,N})}{2n} = \frac{1}{c}\left(x(c) - \frac{x(c)^2}{2}\right)$, where $x(c) = \sum_{v=1}^{\infty} \frac{v^{v-1}(c\mathrm{e}^{-c})^v}{v!}$ and $v = k + l$.

Proof

In order to prove the first two parts we will make use of Theorem 2.3.7, which states the number of cycles. A new link can either connect two components or create at least one cycle. Hence a new link either increases $N - \zeta_{n,N}$ by one or increases $C_{n,N}$ by at least one. Thus it holds that $N \leq n - \zeta_{n,N} + C_{n,N}$.

For $c < 1$ the expected number of cycles is a constant and hence $\mathbb{E}(\zeta_{n,N}) = n - N + O(1)$.

Similarly when $c = 1$ the expected number of cycles is given by $\frac{1}{2}\log(n)$. Therefore it holds that $\mathbb{E}(\zeta_{n,N}) = n - N + O\left(\log(n)\right)$.

Theorem 3.1 implies that the expected number of components of size k,l with $m \geq k + l$ edges is of the order $O(\frac{N^m}{n^{2m-k-l}}) = O\left((\frac{N}{n})^{k+l}\right)$, which is bounded $\forall k$,l. The number of components of size K,L or greater is trivially of order $O(\frac{2n}{K+L})$, where K,$L < n$ can be chosen arbitrarily large. Hence the number of components is similar to the number of spanning trees. As a result it holds that $\mathbb{E}(\zeta_{n,N}) \sim \sum_{k=0}^{n}\sum_{l=0}^{n} \mathbb{E}(\tau_{k,l}) \sim \frac{n^2}{N}\sum_{k=0}^{n}\sum_{l=0}^{n} \frac{k^{l-1}l^{k-1}}{k!l!}(\frac{N}{n}\mathrm{e}^{-\frac{N}{n}})^{k+l}$.

Using $\sum_{k=0}^{v} \frac{k^{v-k-1}(v-k)^{k-1}}{k!(v-k)!} = \frac{2v^{v-2}}{v!}$, it follows that $\lim_{n\to\infty} \frac{\mathbb{E}(\zeta_{n,N})}{2n}$
$= \frac{1}{2c}\sum_{k=0}^{\infty}\sum_{l=0}^{\infty} \frac{k^{l-1}l^{k-1}}{k!l!}(c\mathrm{e}^{-c})^{k+l} = \frac{1}{2c}\sum_{v=1}^{\infty}\sum_{k+l=v} \frac{k^{l-1}l^{k-1}}{k!l!}(c\mathrm{e}^{-c})^{k+l}$
$= \frac{1}{2c}\sum_{v=1}^{\infty}\sum_{k=0}^{v} \frac{k^{v-k-1}(v-k)^{k-1}}{k!(v-k)!}(c\mathrm{e}^{-c})^{v} = \frac{1}{2c}\sum_{v=1}^{\infty}(c\mathrm{e}^{-c})^{v}\sum_{k=0}^{v} \frac{k^{v-k-1}(v-k)^{k-1}}{k!(v-k)!}$
$= \frac{1}{2c}\sum_{v=1}^{\infty}(c\mathrm{e}^{-c})^{v}\frac{2v^{v-2}}{v!} = \frac{1}{c}\sum_{v=1}^{\infty}(c\mathrm{e}^{-c})^{v}\frac{v^{v-2}}{v!} = \frac{1}{c}\left(x(c) - \frac{x(c)^2}{2}\right)$ as in Erdős and Rényi (1960).

□

Appendix B: Theorems in Section 4

Theorem 2.4.1

a) If $p = O(\frac{1}{n^\alpha})$, $\alpha > 1$ it follows that $\mathbb{E}(\text{eff}) \to 0$.

b) If $p = \frac{c}{n}$ it follows that $\mathbb{E}(\text{eff}) \to d \in [1 - \frac{1-\mathrm{e}^{-c}}{c}, 1 - \mathrm{e}^{-c}]$.

c) If $\frac{1}{p} = O(n^\alpha)$, $\alpha < 1$ it follows that $\mathbb{P}(\text{eff} = 1) \to 1$ and therefore $\mathbb{E}(\text{eff}) \to 1$.

Proof

a) We can bound the probability that buyer i trades by the probability that he has at least one link: $\mathbb{P}(\text{buyer}_i \text{ trades}) \leq 1 - (1-p)^n = 1 - (1 - \frac{1}{n^\alpha})^n = 1 - \left((1 - \frac{1}{n^\alpha})^{n^\alpha}\right)^{n^{1-\alpha}} \to 1 - \mathrm{e}^{-n^{1-\alpha}}$ $\to 1 - \mathrm{e}^0 = 0$. Since $\mathbb{P}(\text{buyer}_i \text{ trades}) \to 0$ we have that $\mathbb{E}(\text{eff}) \to 0$.

b) We select buyers one by one and if possible let them trade with a linked and available seller. The i^{th} selected buyer has at least $n - i + 1$ available sellers. Hence a lowerbound is found by considering the probability that this buyer is not linked to at least one of the available $n - i + 1$ sellers. It holds that $1 - (1 - \frac{c}{n})^{n-i+1} \leq \mathbb{P}(i^{th} \text{ selected buyer trades}) \leq 1 - (1 - \frac{c}{n})^n$. Similarly, this leads to $\frac{n+1-\frac{n^\alpha}{c}+\frac{(1-\frac{c}{n^\alpha})^{n+1}}{c}}{n} \leq \mathbb{E}(\text{eff}) \leq \frac{n(1-(1-\frac{c}{n})^n)}{n}$. In the limit it follows that $1 - \frac{1-\mathrm{e}^{-c}}{c} \leq \mathbb{E}(\text{eff}) \leq 1 - \mathrm{e}^{-c}$.

c) Here we make use of the matrix representation. A perfect matching and therefore a maximal number of trades is obtained if there does not exist a block of zeros, where the number of columns plus the number of rows exceeds n. We show that the probability that the maximal number of trades t is lower than n converges to zero.

$\mathbb{P}(t < n) = \mathbb{P}(\exists k, l \text{ block}, k + l > n) \leq \sum_{i=1}^{n} \mathbb{P}(\exists i(n - i + 1) \text{ block})$
$\leq \sum_{i=1}^{n} \binom{n}{i}\binom{n}{n-i+1}(1-p)^{i(n-i+1)} \leq 2\sum_{i=1}^{\lfloor\frac{n+1}{2}\rfloor} \binom{n}{i}\binom{n}{n-i+1}(1-p)^{i(n-i+1)}$. As a result of Claim 2.4.1 it holds that $\mathbb{P}(\text{eff} = 1) \to 1$ and thus $\mathbb{E}(\text{eff}) \to 1$. □

Claim 2.4.1

$\lim_{n\to\infty} \sum_{i=1}^{\lfloor\frac{n+1}{2}\rfloor} \binom{n}{i}\binom{n}{n-i+1}(1-p)^{i(n-i+1)} \to 0$ for $p = O(\frac{1}{n^\alpha})$ and $\alpha < 1$.

Proof

Let us denote $\frac{\mathbb{P}(\exists(i+1)(n-i)\text{ block})}{\mathbb{P}(\exists i(n-i+1)\text{ block})}$ as f_i. We use induction to prove the claim. $f_1 = \frac{1}{1}\cdot\frac{1}{2}n(n-1)(1-p)^{n-2} < 1$ for n large enough. Furthermore $\frac{f_{i+1}}{f_i} = \frac{i}{i+2}\frac{n-i+2}{n-i}(1-p)^2 < 1,\ i \leq \lfloor\frac{n-3}{2}\rfloor$. So $f_i < 1,\ i \leq \lfloor\frac{n-1}{2}\rfloor$ and therefore it holds that $\mathbb{P}(\exists 1\cdot n\text{ block}) \geq \mathbb{P}(\exists i(n-i+1)\text{ block}),\ i \leq \lfloor\frac{n+1}{2}\rfloor$.

Now we can conclude that
$\lim_{n\to\infty}\mathbb{P}(t<n) \leq 2\sum_{i=1}^{\lfloor\frac{n+1}{2}\rfloor}\binom{n}{i}\binom{n}{n-i+1}(1-p)^{i(n-i+1)} \leq 2\sum_{i=1}^{\lfloor\frac{n+1}{2}\rfloor}\binom{n}{1}\binom{n}{n}(1-p)^n$
$\leq 2\lim_{n\to\infty} n\lfloor\frac{n+1}{2}\rfloor(1-p)^n \leq 2\lim_{n\to\infty} n^2(1-p)^n \leq 2\lim_{n\to\infty} n^2\mathrm{e}^{-n^{1-\alpha}} \to 0$ for $p = O(\frac{1}{n^\alpha})$ and $\alpha < 1$. □

Theorem 2.4.2

Although the behaviour of the graph changes abruptly when c passes through 1, the expected efficiency is continuous in c.

Proof

For any given graph G we add a link e. Obviously we have for t the maximal number of trades that $\mathbb{E}\left(t(G+e)\right) - \mathbb{E}\left(t(G)\right) \leq 1$; a removal of the link e can only lead to a decrease of at maximum one trade. So for any number of added links δ we have that $\mathbb{E}\left(t(G+\sum_{i=1}^{\delta} e_i)\right) - \mathbb{E}\left(t(G)\right) \leq \delta$. The expected efficiency as a function of p is simply the weighted sum over all possible graphs. So now we can find a δ for every $\epsilon > 0$ such that $\mathbb{E}\left(\text{eff}(p+\delta)\right) - \mathbb{E}\left(\text{eff}(p)\right) \leq \epsilon$; namely a δ such that $\lim_{n\to\infty}\frac{\delta}{n} \leq \epsilon$. This proves right continuity and similarly the expected efficiency is also left continuous. □

Theorem 2.4.3

Algorithm 2.4.1 produces the set of possible spanning trees of size k,l.

Proof

The maximal in-degree of the vertices is one and all vertices are connected, thus the algorithm only constructs spanning trees.

Next we will prove by reductio ad absurdum that this algorithm produces all the possible spanning trees. Let us suppose a spanning tree T of size k,l exists that is not produced by the algorithm. The set of points in T that are linked to v_1^1 is non-empty and a subset of V^2 and is thus produced by the algorithm. Let D be the maximal distance from node v_1^1. The sets of vertices with distance $0 < d \leq D$ from node v_1^1 must be non-empty and disjoint. If this set is empty there cannot be a point with distance D. If these sets are not disjoint then there is a point from which there are two possible paths to reach v_1^1, which is not possible in a spanning tree. Trivially since T is a spanning tree, all the vertices that do not have distance $0 < d < D$ must have distance D. Thus T must be produced by the algorithm. □

Theorem 2.4.4

A spanning tree with k buyers and l sellers has $k + l - 1$ edges. Every vertex has at least one edge. The remaining $l - 1$ edges are multinomially distributed over the buyers, and the remaining $k - 1$ edges multinomially over the sellers.

Proof

The number of spanning trees with degrees $i_1, ..., i_k$ respectively $j_1, ..., j_l$ can be computed recursively by building such a spanning tree step by step. We select a vertex or component without unfulfilled outgoing links and calculate the number of possible incoming links. The set V^1 has l outgoing links and V^2 has $k - 1$ outgoing links. A selected vertex or component with a root in V^1 can be linked to all $k - 1$ outgoing links of set V^2. A selected vertex or component with a root in V^2 can be linked to $l - 1$ outgoing links; the outgoing link of the root v_1^1 excluded. There have to be at least two vertices with degree one so we can assume without loss of generality that v_1^1 has degree one. After every step there has to be at least one component remaining without any unfulfilled outgoing links. The outgoing link of root v_1^1 has to be fulfilled last, otherwise the remaining vertices or components cannot be added to the same spanning tree. Including symmetric spanning trees for the moment, this results in $(k - 1)!(l - 1)!$

possible spanning trees.

The number of symmetries in a vertex is given by the factorial of the outdegree. We can conclude that the number of different spanning trees of size k,l and degrees $i_1, ..., i_k$ respectively $j_1, ..., j_l$ is given by $\frac{(k-1)!(l-1)!}{(i_1-1)!\cdot...\cdot(i_k-1)!\cdot(j_1-1)!\cdot...\cdot(j_l-1)!}$ and the probability of having those degrees equals $\frac{(k-1)!(l-1)!}{(i_1-1)!\cdot...\cdot(i_k-1)!\cdot(j_1-1)!\cdot...\cdot(j_l-1)!} \cdot \frac{1}{k^{l-1}l^{k-1}}$.

If we only consider the degrees of set V^1, we can calculate the number of spanning trees with degrees $i_1, ..., i_k$ by $\sum_{j_1+...+j_l=k+l-1} \frac{(k-1)!(l-1)!}{(i_1-1)!\cdot...\cdot(i_k-1)!\cdot(j_1-1)!\cdot...\cdot(j_l-1)!} \cdot \frac{1}{k^{l-1}l^{k-1}}$ $= \frac{(l-1)!}{(i_1-1)!\cdot...\cdot(i_k-1)!} \cdot \frac{1}{k^{l-1}} \sum_{j_1+...+j_l=k+l-1} \frac{(k-1)!}{(j_1-1)!\cdot...\cdot(j_l-1)!} \cdot \frac{1}{l^{k-1}} = \frac{(l-1)!}{(i_1-1)!\cdot...\cdot(i_k-1)!} \cdot \frac{1}{k^{l-1}}$ This shows that the remaining edges per set of vertices follow a multinomial distribution, independently of each other. $\square$

Theorem 2.4.5

A lower- and upper bound for the maximum number of trades in a spanning tree with $k+l > 2$ is determined by the number of vertices with a degree of at least 2, $\#V_{\text{degree}>1}$: $LB = \lceil\frac{\#V_{\text{degree}>1}+1}{2}\rceil$ and $UB = \min(k, l, \#V_{\text{degree}>1})$.

Proof

To show the lower bound, we consider the sidebranches until a single path remains. Sidebranches start in vertices with degree larger than two and consist of an isolated path. We consider sidebranches with the root excluded that have vertices with a degree that is maximal two. Vertices in sidebranches are considered only once, and hence a sequence of sidebranches with vertices of a degree maximal two exists. In such a sidebranch the maximal number of trades is equal to $t = \lfloor\frac{\#V^b_{\text{degree}>1}+1}{2}\rfloor \geq \frac{\#V^b_{\text{degree}>1}}{2}$, where V^b denotes the vertices in this branch. By removing such branches one by one only a single path remains. Trivially, the number of possible trades in such a remaining single path is given by $\lceil\frac{\#V^p_{\text{degree}>1}+1}{2}\rceil$, where V^p are the vertices in this path. Hence the total number of possible trades is bounded from below by $\frac{\#V^b_{\text{degree}>1}}{2} + \lceil\frac{\#V^p_{\text{degree}>1}+1}{2}\rceil \geq \lceil\frac{\#V^p_{\text{degree}>1}+1}{2}\rceil$.

Trivially k and l are both an upper bound for the maximum number of trades. Since there are no vertices with degree one linked to each other in a spanning tree with $k+l>2$, every trade includes a vertex with a degree of at least two. Hence $UB = \min(k, l, \#V_{\text{degree}>1})$ is an upper bound for the maximum number of trades. $\square$

Theorem 2.4.6

For $p \sim \frac{c}{n} \iff N \sim cn,\ c \leq 1$ it holds that:
$\sum_{1 \leq i \leq k \leq \infty} \sum_{1 \leq j \leq l \leq \infty} \frac{\mathrm{e}^{-(k+l)}}{i!j!} S(l-1, k-i) S(k-1, l-j) \lceil \frac{k-i+l-j+1}{2} \rceil \leq \mathbb{E}(\text{eff})$
$\leq \sum_{1 \leq i \leq k \leq \infty} \sum_{1 \leq j \leq l \leq \infty} \frac{\mathrm{e}^{-(k+l)}}{i!j!} S(l-1, k-i) S(k-1, l-j) \min(k, l, k-i+l-j)$.

Proof

In a spanning tree every vertex has at least one edge and the remaining edges are multinomially distributed over the vertices. The probability that i out of k vertices do not have any of the remaining $l-1$ edges is given by $\frac{k!}{i!} \frac{S(l-1,k-i)}{k^{l-1}}$, where $S(l-1, k-i)$ is the Stirling number of the second kind. Hence the probability of a spanning tree with k,l vertices where $k-i$ and $l-j$ vertices have a degree larger than 1 equals $\frac{k^{l-1} l^{k-1}}{k!l!} \mathrm{e}^{-(k+l)} \frac{k!}{i!} \frac{S(l-1,k-i)}{k^{l-1}} \frac{l!}{j!} \frac{S(k-1,l-j)}{l^{k-1}}$
$= \frac{\mathrm{e}^{-(k+l)}}{i!j!} S(l-1, k-i) S(k-1, l-j)$.

The result is obtained by multiplying by the bounds on the expected maximal efficiency per spanning tree and summing over all sizes of the spanning tree and the number of vertices with a degree equal to 1: $\sum_{1 \leq i \leq k \leq \infty} \sum_{1 \leq j \leq l \leq \infty} \frac{\mathrm{e}^{-(k+l)}}{i!j!} S(l-1, k-i) S(k-1, l-j) \lceil \frac{k-i+l-j+1}{2} \rceil$
$\leq \mathbb{E}(\text{eff}) \leq \sum_{1 \leq i \leq k \leq \infty} \sum_{1 \leq j \leq l \leq \infty} \frac{\mathrm{e}^{-(k+l)}}{i!j!} S(l-1, k-i) S(k-1, l-j) \min(k, l, k-i+l-j)$.

$\square$