C.Ö. Sayin

The Friendship Paradox: Higher Levels

Bachelor thesis

11 July 2023

Thesis supervisors: prof.dr. W.T.F. Den Hollander
dr. R.S. Hazra

Leiden University
Mathematical Institute

Contents

1 The Friendship Paradox

In his 1991 paper in the *American Journal of Sociology* [1], American sociologist Scott Feld published his discovery that, on average, your friends are more popular than you are. This phenomenon turns out to be true for any social network and, in its graph-theoretical formulation, graph. This so-called Friendship Paradox was also investigated in the 2021 paper by G. Cantwell, A. Kirkley and M. Newman [2]. In his 2022 Bachelor thesis on the Friendship Paradox [3], P. MacDonald proposed, as a suggestion for further research, investigating whether the phenomenon occurs for higher level friendships, such as when looking at the friends of the friends of your friends. This Bachelor thesis aims to do so.

This first chapter provides an overview of the classical Friendship Paradox as discussed in the Bachelor thesis of P. MacDonald.

1.1 Introduction

The Friendship Paradox can be understood as follows. We consider a group of n people with mutual friendships. For each person we can compute the difference between the average number of friends of the friends of this person and the number of friends of this person. The average of this difference over all group members turns out to be non-negative. In other words, the average person in the group is less popular than the friends of this person, which is the statement of the classical Friendship Paradox. The Friendship Paradox has been applied to reduction of the spread of viruses [4] and efficient polling methods [5].

1.2 Preliminaries

The formulation and proof of the classical Friendship Paradox requires some mathematical preliminaries. These will be reviewed in this section.

1.2.1 Graphs

In essence, the Friendship Paradox is a graph-theoretical theorem. A *graph* is given by a set of vertices $V = \{1, 2, \ldots, n\}$ and a set of edges $E \subseteq V^2$. In this bachelor thesis, only *undirected* and *simple* graphs will be considered.

Definition 1.1. *A graph G is called undirected if its edges have no orientation, i.e. $(i, j) = (j, i)$ for $i, j \in V$.*

Definition 1.2. *An undirected graph G is called simple if it does not contain self-loops or multiple edges, i.e. $(i, i) \notin E$ for all $i \in V$ and E is not a multiset.*

In terms of social networks,

- undirectedness corresponds to the friendships being mutual;
- simplicity corresponds to individuals not being friends with themselves and friendships being represented by one edge only.

Initially, we will restrict the analysis to connected graphs, i.e. graphs with a path between any two vertices.

1.2.2 The adjacency matrix

A graph G can be represented by its *adjacency matrix* A. For a graph with no multiple edges, hence for simple graphs, A is given by

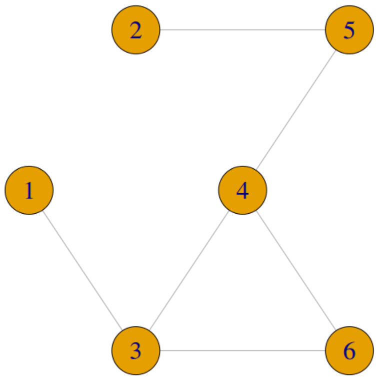
$$A_{ij} = \begin{cases} 1 & \text{if } (i,j) \in E, \\ 0 & \text{otherwise.} \end{cases}$$

For a vertex i, its *degree* is its number of neighbours, i.e.

$$d_i = \sum_{j=1}^{n} A_{ij}.$$

Note that for undirected graphs, the adjacency matrix is symmetric since $A_{ij} = A_{ji}$ for all $i,j \in V$. Also note that for a connected graph, all vertices have degree at least 1.

Example 1.3. *The following graph G with $n = 6$ is undirected, simple and connected.*



It is represented by the symmetric adjacency matrix

$$A = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \end{bmatrix}.$$

1.3 Formulation of the Classical Friendship Paradox

The classical Friendship Paradox can now be formulated.

Consider an undirected, simple and connected graph G. For a vertex i, its *friendship bias* Δ_i is defined as

$$\Delta_i := \frac{1}{d_i} \sum_{j=1}^{n} A_{ij} d_j - d_i,$$

i.e. the difference between the average degree of its neighbours and its own degree. The Friendship Paradox states that the average of the friendship biases is non-negative, and is zero if and only if all degrees are equal.

Theorem 1.4 [3, Theorem 1]. *If G is an undirected, simple and connected graph, then*

$$\Delta := \frac{1}{n}\sum_{i=1}^{n}\Delta_i \geq 0,$$

with equality if and only if all degrees are equal.

Proof. Compute

$$\begin{aligned}
\Delta &= \frac{1}{n}\sum_{i=1}^{n}\Delta_i \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{d_i}\sum_{j=1}^{n}A_{ij}d_j - d_i\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{d_i}\sum_{j=1}^{n}A_{ij}d_j - \sum_{j=1}^{n}A_{ij}\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}A_{ij}\left(\frac{d_j}{d_i} - 1\right) \\
&= \frac{1}{2n}\sum_{i=1}^{n}\sum_{j=1}^{n}A_{ij}\left(\frac{d_j}{d_i} + \frac{d_i}{d_j} - 2\right) \\
&= \frac{1}{2n}\sum_{i=1}^{n}\sum_{j=1}^{n}A_{ij}\left(\sqrt{\frac{d_j}{d_i}} - \sqrt{\frac{d_i}{d_j}}\right)^2 \\
&\geq 0,
\end{aligned}$$

with equality if and only if all degrees are equal. □

Given that not all individuals have the same number of friends in a social network, this can be interpreted as saying that on average, your friends are more popular than you.

Example 1.5. *For the graph in Example 1.3 we have*

$$(\Delta_i)_{i\in\{1,\dots,6\}} = (2, 1, -1, -\frac{2}{3}, 0, 1)$$

and, in agreement with the Friendship Paradox,

$$\Delta = \frac{1}{6}\sum_{i=1}^{6}\Delta_i = \frac{7}{18} > 0$$

The Friendship Paradox can also be stated in terms of degree expectations as follows.

Theorem 1.6 [3, Theorem 2]. *Let $G = (V, E)$ be an undirected, simple and connected graph. Let $\bar{U}$ be a uniformly chosen vertex from V, and let $e = (U, W)$ be a uniformly chosen edge from E. Then $\mathbb{E}[d_U] \geq \mathbb{E}[d_{\bar{U}}] \geq 0$. The interpretation of this inequality is that the number of friends of an individual in a randomy chosen friendship is greater than or equal to that of a randomly chosen individual.*

For the proof of Theorem 1.6, refer to [3, Theorem 2].

1.4 Outlook

The theory in this section will be expanded upon in the following sections. In particular, in Chapter 2 we will consider the popularity of friends at higher levels than your direct friends, e.g. the friends of your friends, and deal with graphs with isolated vertices. The Erdős–Rényi random graph will be introduced in Chapter 3, and for this random graph some expectations will be computed. Finally, in Chapter 4 we will investigate the consequences of letting the level of friendship or the amount of individuals in the network tend to infinity.

2 The Higher Level Friendship Paradox

The Friendship Paradox says that, in a social network, the average difference between the average number of friends of an individual's friends and the number of friends of that individual is non-negative. We seek to compare average numbers of friends at higher levels, i.e. friends that are two or more steps away from the initial individual.

2.1 Motivation

The Friendship Paradox states that on average, your direct friends are more popular than you are (given not all individuals in your friend group have the same number of friends). This gives rise to the question how popular the *friends of your friends* are, or even your friends at higher levels, and how these popularities compare. For example, are the friends of your friends (your second-level friends) more or less popular than *their* friends (your third-level friends)?

2.2 Definitions

In order to extend our analysis to higher levels of friendship we introduce some new quantities.

Let G be an undirected, simple and connected graph. (Later it will be shown that the connectedness assumption can be dropped without affecting our analysis.) Fix $i \in V$ and let $k, l \in \mathbb{N}$. Then the number of friends of friends of individual i at the k-th level is given by $\sum_{j=1}^{n}(A^k)_{ij}d_j$, where A^k is the k-th power of the adjacency matrix A. Indeed, for an individual j, its degree d_j contributes to this sum if and only if $(A^k)_{ij} = 1$, i.e. if j is k edges away from i, so that j is a k-th level friend of i.

Now, averaging over the friends of i yields the quantity

$$\frac{1}{d_i}\sum_{j=1}^{n}(A^k)_{ij}d_j$$

as the average number of friends of friends at the k-th level for $k \geq 1$. This can be interpreted as the relative popularity of the k-th level friends of i. Note that, for $k \geq 2$, i itself may be included in the k-th level friends.

Thus, we define the average number of friends of friends at the k-th level as

$$\chi_i^{(k)} := \begin{cases} \frac{1}{d_i}\sum_{j=1}^{n}(A^k)_{ij}d_j & \text{if } k \geq 1, \\ d_i & \text{if } k = 0, \end{cases} \tag{1}$$

with the average over all individuals given by

$$\chi_{[n]}^{(k)} := \frac{1}{n}\sum_{i=1}^{n}\chi_i^{(k)}.$$

In order to compare these averages, or popularities, between levels k and l, we introduce the differences

$$\Delta_i^{(k,l)} := \chi_i^{(k)} - \chi_i^{(l)}$$

for $0 \leq l < k$, where

$$\Delta_{[n]}^{(k,l)} := \frac{1}{n}\sum_{i=1}^{n}\Delta_i^{(k,l)}.$$

If $k > 0$, then we write

$$\Delta_i^{(k)} := \Delta_i^{(k,k-1)} \tag{2}$$

and

$$\Delta_{[n]}^{(k)} := \Delta_{[n]}^{(k,k-1)}.$$

Lemma 2.1. *Suppose that $0 \leq l < k$. For every undirected, simple and connected graph G,*

$$\Delta_i^{(k,l)} = \sum_{m=l+1}^{k} \Delta_i^{(m)}$$

and

$$\Delta_{[n]}^{(k,l)} = \sum_{m=l+1}^{k} \Delta_{[n]}^{(m)}$$

Proof. This follows from the fact that $\Delta_i^{(k,l)}$ and $\Delta_{[n]}^{(k,l)}$ are telescoping series. □

2.3 Formulation of the Higher Level Friendship Paradox

Recall the statement of the Friendship Paradox in Theorem 1.4. The Friendship Paradox can now be expressed as follows.

Theorem 2.2. *If G is an undirected, simple and connected graph, then*

$$\Delta_{[n]}^{(1)} = \frac{1}{n} \sum_{i=1}^{n} \Delta_i^{(1)} \geq 0,$$

with equality if and only if all degrees are equal.

It turns out that the Friendship Paradox can also be extended to higher levels as follows.

Theorem 2.3. *Let $k \geq 2$. If G is an undirected, simple and connected graph, then*

$$\Delta_i^{(k)} \geq 0$$

and hence

$$\Delta_{[n]}^{(k)} = \frac{1}{n} \sum_{i=1}^{n} \Delta_i^{(k)} \geq 0.$$

Proof. Since $d_j \geq 1$ for all j, we have

$$\begin{aligned}
\Delta_i^{(k)} &= \chi_i^{(k)} - \chi_i^{(k-1)} \\
&= \frac{1}{d_i}\left(\sum_{j=1}^{n}(A^k)_{ij}d_j - \sum_{j=1}^{n}(A^{k-1})_{ij}d_j\right) \\
&= \frac{1}{d_i}\left(\sum_{j=1}^{n}\sum_{l=1}^{n}(A^{k-1})_{il}A_{lj}d_j - \sum_{j=1}^{n}(A^{k-1})_{ij}d_j\right) \\
&= \frac{1}{d_i}\sum_{l=1}^{n}(A^{k-1})_{il}\left(\sum_{j=1}^{n}A_{lj}d_j - d_l\right) \\
&= \frac{1}{d_i}\sum_{l=1}^{n}(A^{k-1})_{il}\sum_{j=1}^{n}A_{lj}\left(d_j - 1\right) \\
&\geq 0.
\end{aligned}$$

□

For $k = 2$, this can be interpreted as saying that, on average, the friends of your friends are more popular than your friends.

2.4 Graphs with isolated vertices

So far we have only considered connected graphs. These graphs have no isolated vertices, corresponding to social networks in which every individual has at least one friend. In real-life networks and in random graphs, however, these can occur. If i is such an isolated vertex then $d_i = 0$, so that the expression

$$\frac{1}{d_i}\sum_{j=1}^{n}(A^k)_{ij}d_j$$

(where $k > 0$) makes no sense. To fix this, we define, for $k > 0$,

$$\Delta_i^{(k)*} := \begin{cases} \Delta_i^{(k)} & \text{if } d_i > 0, \\ 1 & \text{if } d_i = 0, \end{cases}$$

where $\Delta_i^{(k)}$ is as in (2), and

$$\Delta_{[n]}^{(k)*} = \frac{1}{n}\sum_{i=1}^{n}\Delta_i^{(k)*}.$$

With this alteration, the (extended) Friendship Paradox still holds.

Theorem 2.4. *Let $k > 0$. If G is an undirected and simple graph, then*

$$\Delta_{[n]}^{(k)*} = \frac{1}{n}\sum_{i=1}^{n}\Delta_i^{(k)*} \geq 0.$$

Proof. First suppose that $k = 1$. Compute

$$\begin{aligned}
\Delta_{[n]}^{(1)*} &= \frac{1}{n}\sum_{i=1}^{n}\left[\Delta_i^{(1)}\mathbb{1}_{\{d_i>0\}} + \mathbb{1}_{\{d_i=0\}}\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\left[\left(\frac{1}{d_i}\sum_{j=1}^{n}A_{ij}d_j - d_i\right)\mathbb{1}_{\{d_i>0\}} + \mathbb{1}_{\{d_i=0\}}\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\left[\sum_{j=1}^{n}A_{ij}\left(\frac{d_j}{d_i}-1\right)\mathbb{1}_{\{d_i>0\}} + \mathbb{1}_{\{d_i=0\}}\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\left[\sum_{j=1}^{n}A_{ij}\left(\frac{d_j}{d_i}-1\right)\mathbb{1}_{\{d_i>0\}}\mathbb{1}_{\{d_j>0\}} + \sum_{j=1}^{n}A_{ij}\left(\frac{d_j}{d_i}-1\right)\mathbb{1}_{\{d_i>0\}}\mathbb{1}_{\{d_j=0\}} + \mathbb{1}_{\{d_i=0\}}\right],
\end{aligned} \tag{3}$$

since $\mathbb{1}_{\{d_j>0\}} + \mathbb{1}_{\{d_j=0\}} = 1$ for all j.

For the first term in (3),

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}A_{ij}\left(\frac{d_j}{d_i}-1\right)\mathbb{1}_{\{d_i>0\}}\mathbb{1}_{\{d_j>0\}} &= \frac{1}{2n}\sum_{i=1}^{n}\sum_{j=1}^{n}A_{ij}\left(\frac{d_j}{d_i}+\frac{d_i}{d_j}-2\right)\mathbb{1}_{\{d_i>0\}}\mathbb{1}_{\{d_j>0\}} \\
&= \frac{1}{2n}\sum_{i=1}^{n}\sum_{j=1}^{n}A_{ij}\left(\sqrt{\frac{d_j}{d_i}}-\sqrt{\frac{d_i}{d_j}}\right)^2\mathbb{1}_{\{d_i>0\}}\mathbb{1}_{\{d_j>0\}} \\
&\geq 0.
\end{aligned}$$

For the second term in (3),

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}A_{ij}\left(\frac{d_j}{d_i}-1\right)\mathbb{1}_{\{d_i>0\}}\mathbb{1}_{\{d_j=0\}} = -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}A_{ij}\mathbb{1}_{\{d_i>0\}}\mathbb{1}_{\{d_j=0\}} = 0,$$

since $d_j = 0$ if and only if $A_{ji} = A_{ij} = 0$ for all i.

For the third term in (3),

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{\{d_i=0\}} \geq 0.$$

Hence,

$$\Delta_{[n]}^{(1)*} \geq 0,$$

i.e. the statement holds for $k = 1$.

Now suppose that $k \geq 2$. For all i,

$$\begin{aligned}
\Delta_i^{(k)*} &= \Delta_i^{(k)} \mathbb{1}_{\{d_i>0\}} + \mathbb{1}_{\{d_i=0\}} \\
&= \frac{1}{d_i}\left(\sum_{j=1}^{n}(A^k)_{ij}d_j - \sum_{j=1}^{n}(A^{k-1})_{ij}d_j\right)\mathbb{1}_{\{d_i>0\}} + \mathbb{1}_{\{d_i=0\}} \\
&= \frac{1}{d_i}\sum_{l=1}^{n}(A^{k-1})_{il}\sum_{j=1}^{n}A_{lj}\,(d_j-1)\,\mathbb{1}_{\{d_i>0\}} + \mathbb{1}_{\{d_i=0\}} \\
&= \frac{1}{d_i}\sum_{l=1}^{n}(A^{k-1})_{il}\sum_{j=1}^{n}A_{lj}\,(d_j-1)\,\mathbb{1}_{\{d_i>0\}}\left(\mathbb{1}_{\{d_j>0\}} + \mathbb{1}_{\{d_j=0\}}\right) + \mathbb{1}_{\{d_i=0\}}. \qquad (4)
\end{aligned}$$

For the first term in (4),

$$\frac{1}{d_i}\sum_{l=1}^{n}(A^{k-1})_{il}\sum_{j=1}^{n}A_{lj}\,(d_j-1)\,\mathbb{1}_{\{d_i>0\}}\mathbb{1}_{\{d_j>0\}} \geq 0.$$

For the second term in (4),

$$\frac{1}{d_i}\sum_{l=1}^{n}(A^{k-1})_{il}\sum_{j=1}^{n}A_{lj}\,(d_j-1)\,\mathbb{1}_{\{d_i>0\}}\mathbb{1}_{\{d_j=0\}} = -\frac{1}{d_i}\sum_{l=1}^{n}(A^{k-1})_{il}\sum_{j=1}^{n}A_{lj}\mathbb{1}_{\{d_i>0\}}\mathbb{1}_{\{d_j=0\}} = 0,$$

since $d_j = 0$ if and only if $A_{jl} = A_{lj} = 0$ for all l.

For the third term in (4),

$$\mathbb{1}_{\{d_i=0\}} \geq 0.$$

It follows that $\Delta_i^{(k)*} \geq 0$. Hence,

$$\Delta_{[n]}^{(k)*} = \frac{1}{n}\sum_{i=1}^{n}\Delta_i^{(k)*} \geq 0,$$

i.e. the statement holds for $k \geq 2$ as well. □

2.5 Chapter conclusion

In this chapter, we introduced some new quantities, then extended the Friendship Paradox to higher levels of friendship in Theorem 2.3. We have also dealt with isolated vertices and concluded that the extended Friendship Paradox still holds by Theorem 2.4.

3 The Erdős–Rényi Random Graph

We now turn our attention to random graph models. Such models can be useful for simulating real-life networks. In this chapter we investigate the properties of the Erdős–Rényi random graph.

3.1 Definition

Definition 3.1. *An Erdős–Rényi random graph $ER_n(p)$ with parameters $n \in \mathbb{N}$ and $p \in [0,1]$ is an undirected and simple graph generated as follows: for vertex set $V = \{1, 2, \ldots, n\}$, include each possible edge with probability p, independently of the other edges.*

It follows that for each vertex i we have

$$A_{ij} \sim \mathrm{Ber}(p) \text{ for all } j \neq i,$$

where $\mathrm{Ber}(p)$ is a Bernoulli random variable with parameter p.

Since there are $n-1$ possible edges per vertex, we have

$$d_i = \sum_{j=1}^{n} A_{ij} \sim \mathrm{Bin}(n-1, p),$$

where $Bin(n-1, p)$ is a Binomial random variable with parameters $n-1$ and p.

It follows that $\mathbb{E}[d_i] = (n-1)p$. This degree expectation diverges as $n \to \infty$ for $p \in [0,1]$, so that $\mathrm{ER}_n(p)$ is dense.

We write

$$p_k := \mathbb{P}(d_i = k).$$

3.2 Isolated vertices

In a realisation of $\mathrm{ER}_n(p)$, isolated vertices can occur, especially for low n. Indeed, the probability for a vertex i to be isolated is

$$p_0 = (1-p)^{n-1},$$

which vanishes as $n \to \infty$.

We can introduce sparsity in $\mathrm{ER}_n(p)$ by choosing $p = \frac{\lambda}{n-1}$ for some $\lambda \in (0, \infty)$. Then

$$d_i \sim \mathrm{Bin}(n-1, p) \to \mathrm{Poisson}(\lambda) \text{ as } n \to \infty \text{ [3, Theorem 3]}.$$

Furthermore, we have

$$p_0 = \left(1 - \frac{\lambda}{n-1}\right)^{n-1} \to e^{-\lambda} \text{ as } n \to \infty,$$

i.e. the probability that isolated vertices occur does not vanish as $n \to \infty$.

Note that for this choice of p we have, for all i, $\mathbb{E}[d_i] = (n-1)p = \lambda$. Hence, the degree expectation converges as $n \to \infty$, so that $\mathrm{ER}_n\left(\frac{\lambda}{n}\right)$ is sparse, i.e. the degrees are typically finite.

3.3 Expectations

Since the extended Friendship Paradox in Theorem 2.4 does not give us quantitative estimates for the $\Delta_i^{(k)*}$ or $\Delta_{[n]}^{(k)*}$, we seek qualitative bounds. Hence, we consider the Erdős–Rényi random graph with $p \in [0,1]$ and set out to compute $\mathbb{E}\left[\Delta_i^{(1)*}\right]$, $\mathbb{E}\left[\Delta_i^{(2)*}\right]$ and their difference. In order to do so we first introduce some lemmas.

Lemma 3.2.

$$\mathbb{E}\left[d_i \mathbb{1}_{\{d_i>0\}}\right] = (n-1)p.$$

Proof. Compute

$$\begin{aligned}
\mathbb{E}\left[d_i \mathbb{1}_{\{d_i>0\}}\right] &= \mathbb{E}\left[d_i \mathbb{1}_{\{d_i>0\}}\right] + 0p_0 \\
&= \mathbb{E}\left[d_i \mathbb{1}_{\{d_i>0\}}\right] + \mathbb{E}\left[d_i \mathbb{1}_{\{d_i=0\}}\right] \\
&= \mathbb{E}\left[d_i\right] \\
&= (n-1)p.
\end{aligned}$$

□

Lemma 3.3.

$$\mathbb{E}\left[\chi_i^{(1)} \mathbb{1}_{\{d_i>0\}}\right] = [1+(n-2)p][1-p_0],$$

where $\chi_i^{(1)}$ is as in (1).

Proof. Compute

$$\begin{aligned}
\mathbb{E}\left[\chi_i^{(1)} \mathbb{1}_{\{d_i>0\}}\right] &= \mathbb{E}\left[\frac{1}{d_i}\sum_{j=1}^{n} A_{ij} d_j \mathbb{1}_{\{d_i>0\}}\right] \\
&= \mathbb{E}\left[\frac{1}{d_i}\sum_{j=1}^{n}\sum_{h=1}^{n} A_{ij}A_{jh} \mathbb{1}_{\{d_i>0\}}\right] \\
&= \sum_{k=1}^{n-1}\frac{p_k}{k}\sum_{j=1}^{n}\sum_{h=1}^{n} \mathbb{E}\left[A_{ij}A_{jh} | d_i = k\right].
\end{aligned}$$

In this sum we distinguish between the cases $h = i$ and $h \neq i$ as follows:

- $h = i$:

$$\begin{aligned}
\sum_{k=1}^{n-1}\frac{p_k}{k}\sum_{\substack{j=1\\ j\neq i}}^{n} \mathbb{E}\left[A_{ij}A_{ji} | d_i = k\right] &= \sum_{k=1}^{n-1}\frac{p_k}{k}\sum_{\substack{j=1\\ j\neq i}}^{n} \mathbb{E}\left[A_{ij} | d_i = k\right] \\
&= \sum_{k=1}^{n-1}\frac{p_k}{k}\mathbb{E}\left[d_i | d_i = k\right] \\
&= \sum_{k=1}^{n-1} p_k \\
&= 1 - p_0.
\end{aligned}$$

- $h \neq i$:

$$\sum_{k=1}^{n-1} \frac{p_k}{k} \sum_{\substack{j=1 \\ j\neq i}}^{n} \sum_{\substack{h=1 \\ h\neq i}}^{n} \mathbb{E}\left[A_{ij}A_{jh} | d_i = k\right] = \sum_{k=1}^{n-1} \frac{p_k}{k} \sum_{\substack{j=1 \\ j\neq i}}^{n} \sum_{\substack{h=1 \\ h\notin\{i,j\}}}^{n} \mathbb{E}\left[A_{ij}A_{jh} | d_i = k\right]$$
$$= (n-2)p \sum_{k=1}^{n-1} \frac{p_k}{k} \sum_{\substack{j=1 \\ j\neq i}}^{n} \mathbb{E}\left[A_{ij} | d_i = k\right]$$
$$= (n-2)p \sum_{k=1}^{n-1} p_k$$
$$= (n-2)p[1-p_0].$$

Addition of these terms yields the expectation. □

Lemma 3.4.

$$\mathbb{E}\left[\chi_i^{(2)} \mathbb{1}_{\{d_i>0\}}\right] = n^2p^2 - n^2p_0p^2 - 4np^2 + 5np_0p^2 + 2np - 6p_0p^2 + 5p^2 + 2p_0p - 4p,$$

where $\chi_i^{(2)}$ is as in (1).

Proof. Compute

$$\mathbb{E}\left[\chi_i^{(2)} \mathbb{1}_{\{d_i>0\}}\right] = \mathbb{E}\left[\frac{1}{d_i} \sum_{j=1}^{n} (A^2)_{ij} d_j \mathbb{1}_{\{d_i>0\}}\right]$$
$$= \mathbb{E}\left[\frac{1}{d_i} \sum_{j=1}^{n} \sum_{l=1}^{n} A_{il}A_{lj} d_j \mathbb{1}_{\{d_i>0\}}\right]$$
$$= \mathbb{E}\left[\frac{1}{d_i} \sum_{j=1}^{n} \sum_{l=1}^{n} \sum_{h=1}^{n} A_{il}A_{lj}A_{jh} \mathbb{1}_{\{d_i>0\}}\right]$$
$$= \sum_{k=1}^{n-1} \frac{p_k}{k} \sum_{j=1}^{n} \sum_{l=1}^{n} \sum_{h=1}^{n} \mathbb{E}\left[A_{il}A_{lj}A_{jh} | d_i = k\right].$$

In order to deal with the dependency between edges sharing vertices in $\{i, j, l, h\}$, we distinguish between the cases $h = i$; $h = l, j = i$; $h = l, j \neq i$; $h \notin \{i, l\}, j = i$; and $h \notin \{i, l\}, j \neq i$ in this sum as follows:

- $h = i$:

$$\begin{aligned}
\sum_{k=1}^{n-1} \frac{p_k}{k} \sum_{j=1}^{n} \sum_{l=1}^{n} \mathbb{E}\left[A_{il} A_{lj} A_{ji} | d_i = k\right] &= \sum_{k=1}^{n-1} \frac{p_k}{k} \sum_{\substack{j=1 \\ j \neq i}}^{n} \sum_{\substack{l=1 \\ l \notin \{i,j\}}}^{n} \mathbb{E}\left[A_{il} A_{lj} A_{ji} | d_i = k\right] \\
&= p \sum_{k=1}^{n-1} \frac{p_k}{k} \sum_{\substack{j=1 \\ j \neq i}}^{n} \sum_{\substack{l=1 \\ l \notin \{i,j\}}}^{n} \mathbb{E}\left[A_{il} A_{ji} | d_i = k\right] \\
&= p \sum_{k=1}^{n-1} \frac{p_k}{k} \sum_{\substack{j=1 \\ j \neq i}}^{n} \mathbb{E}\left[(d_i - A_{ij}) A_{ij} | d_i = k\right] \\
&= p \left(\sum_{k=1}^{n-1} \frac{p_k}{k} \sum_{\substack{j=1 \\ j \neq i}}^{n} \mathbb{E}\left[k A_{ij} | d_i = k\right] - \sum_{k=1}^{n-1} \frac{p_k}{k} \sum_{\substack{j=1 \\ j \neq i}}^{n} \mathbb{E}\left[A_{ij} | d_i = k\right] \right) \\
&= p \left(\sum_{k=1}^{n-1} p_k k - \sum_{k=1}^{n-1} p_k \right) \\
&= p \left((n-1)p - (1 - p_0) \right) \\
&= (n-1)p^2 - (1-p_0)p.
\end{aligned}$$

- $h = l, j = i$

$$\begin{aligned}
\sum_{k=1}^{n-1} \frac{p_k}{k} \sum_{l=1}^{n} \mathbb{E}\left[A_{il} A_{li} A_{il} | d_i = k\right] &= \sum_{k=1}^{n-1} \frac{p_k}{k} \sum_{\substack{l=1 \\ l \neq i}}^{n} \mathbb{E}\left[A_{il} | d_i = k\right] \\
&= \sum_{k=1}^{n-1} p_k \\
&= 1 - p_0.
\end{aligned}$$

- $h = l, j \neq i$:

$$\sum_{k=1}^{n-1} \frac{p_k}{k} \sum_{\substack{j=1 \\ j\neq i}}^{n} \sum_{l=1}^{n} \mathbb{E}\left[A_{il}A_{lj}A_{jl} | d_i = k\right] = \sum_{k=1}^{n-1} \frac{p_k}{k} \sum_{\substack{j=1 \\ j\neq i}}^{n} \sum_{\substack{l=1 \\ l\notin\{i,j\}}}^{n} \mathbb{E}\left[A_{il}A_{lj} | d_i = k\right]$$
$$= (n-1)p \sum_{k=1}^{n-1} \frac{p_k}{k} \sum_{\substack{l=1 \\ l\notin\{i,j\}}}^{n} \mathbb{E}\left[A_{il} | d_i = k\right]$$
$$= (n-1)p \sum_{k=1}^{n-1} \frac{p_k}{k} \mathbb{E}\left[d_i - A_{ij} | d_i = k\right]$$
$$= (n-1)p \left(\sum_{k=1}^{n-1} p_k - \sum_{k=1}^{n-1} \frac{p_k}{k} \mathbb{E}\left[A_{ij} | d_i = k\right] \right)$$
$$= (n-1)p \left(1 - \sum_{k=1}^{n-1} \frac{p_k}{k} \mathbb{E}\left[A_{ij} | d_i = k\right] \right). \tag{5}$$

We compute the second term between brackets separately as follows.

$$\sum_{k=1}^{n-1} \frac{p_k}{k} \mathbb{E}\left[A_{ij} | d_i = k\right] = \sum_{k=1}^{n-1} \frac{p_k}{k} \mathbb{P}\left(A_{ij} = 1 | d_i = k\right)$$
$$= \sum_{k=1}^{n-1} \frac{1}{k} \mathbb{P}\left(A_{ij} = 1, d_i = k\right)$$
$$= p \sum_{k=1}^{n-1} \frac{1}{k} \mathbb{P}\left(d_i = k | A_{ij} = 1\right)$$
$$= p \sum_{k=1}^{n-1} \frac{1}{k} \mathbb{P}\left(\mathrm{Bin}(n-2, p) = k-1\right)$$
$$= p \sum_{k=1}^{n-1} \frac{1}{k} \frac{(n-2)!}{(k-1)!(n-k-1)!} p^{k-1} q^{n-k-1}$$
$$= \frac{1}{n-1} \sum_{k=1}^{n-1} \frac{(n-1)!}{k!(n-k-1)!} p^k q^{n-k-1}$$
$$= \frac{1}{n-1} \sum_{k=1}^{n-1} \mathbb{P}\left(\mathrm{Bin}(n-1, p) = k\right)$$
$$= \frac{1}{n-1}(1 - p_0). \tag{6}$$

Plugging (6) into (5) we get

$$\sum_{k=1}^{n-1} \frac{p_k}{k} \sum_{\substack{j=1 \\ j\neq i}}^{n} \sum_{l=1}^{n} \mathbb{E}\left[A_{il}A_{lj}A_{jl} | d_i = k\right] = (n-1)p \left(1 - \frac{1}{n-1}(1 - p_0) \right)$$
$$= (n-2)p + pp_0.$$

- $h \notin \{i,l\}, j = i$:

$$\begin{aligned}
\sum_{k=1}^{n-1} \frac{p_k}{k} \sum_{l=1}^{n} \sum_{\substack{h=1 \\ h \notin \{i,l\}}}^{n} \mathbb{E}\left[A_{il}A_{li}A_{ih} | d_i = k\right] &= \sum_{k=1}^{n-1} \frac{p_k}{k} \sum_{\substack{l=1 \\ l \neq i}}^{n} \sum_{\substack{h=1 \\ h \notin \{i,l\}}}^{n} \mathbb{E}\left[A_{il}A_{ih} | d_i = k\right] \\
&= \sum_{k=1}^{n-1} \frac{p_k}{k} \sum_{\substack{l=1 \\ l \neq i}}^{n} \mathbb{E}\left[A_{il}(d_i - A_{il}) | d_i = k\right] \\
&= \sum_{k=1}^{n-1} \frac{p_k}{k} \sum_{\substack{l=1 \\ l \neq i}}^{n} \mathbb{E}\left[kA_{il} | d_i = k\right] - \sum_{k=1}^{n-1} \frac{p_k}{k} \sum_{\substack{l=1 \\ l \neq i}}^{n} \mathbb{E}\left[A_{il} | d_i = k\right] \\
&= \sum_{k=1}^{n-1} p_k k - \sum_{k=1}^{n-1} p_k \\
&= (n-1)p - (1-p_0) \\
&= (n-1)p - 1 + p_0.
\end{aligned}$$

- $h \notin \{i,l\}, j \neq i$:

$$\begin{aligned}
\sum_{k=1}^{n-1} \frac{p_k}{k} \sum_{\substack{j=1 \\ j \neq i}}^{n} \sum_{l=1}^{n} \sum_{\substack{h=1 \\ h \notin \{i,l\}}}^{n} \mathbb{E}\left[A_{il}A_{lj}A_{jh} | d_i = k\right] &= \sum_{k=1}^{n-1} \frac{p_k}{k} \sum_{\substack{j=1 \\ j \neq i}}^{n} \sum_{\substack{l=1 \\ l \notin \{i,j\}}}^{n} \sum_{\substack{h=1 \\ h \notin \{i,j,l\}}}^{n} \mathbb{E}\left[A_{il}A_{lj}A_{jh} | d_i = k\right] \\
&= p^2(n-2)(n-3) \sum_{k=1}^{n-1} \frac{p_k}{k} \sum_{\substack{j=1 \\ j \neq i}}^{n} \mathbb{E}\left[A_{il} | d_i = k\right] \\
&= p^2(n-2)(n-3) \sum_{k=1}^{n-1} p_k \\
&= p^2(n-2)(n-3)(1-p_0).
\end{aligned}$$

Addition of these terms yields the expectation. □

From the above lemmas we conclude the following.

Theorem 3.5.

$$\mathbb{E}\left[\Delta_i^{(1)*}\right] = 1 - (n-2)pp_0 - p.$$

Proof. This follows from the fact that

$$\begin{aligned}
\mathbb{E}\left[\Delta_i^{(1)*}\right] &= \mathbb{E}\left[\Delta_i^{(1)} 1_{\{d_i > 0\}} + 1 \cdot \mathbb{1}_{\{d_i = 0\}}\right] \\
&= \mathbb{E}\left[\chi_i^{(1)} \mathbb{1}_{\{d_i > 0\}}\right] - \mathbb{E}\left[d_i \mathbb{1}_{\{d_i > 0\}}\right] + p_0.
\end{aligned}$$

Using Lemmas 3.3 and 3.2 yields the result. □

Theorem 3.6.

$$\mathbb{E}\left[\Delta_i^{(2)*}\right] = n^2p^2(1-p_0) + 5np^2p_0 - 4np^2 + np(1+p_0) - 6p^2p_0 + 5p^2 - 2p + 2p_0 - 1.$$

Proof. This follows from the fact that

$$\begin{aligned}\mathbb{E}\left[\Delta_i^{(2)*}\right] &= \mathbb{E}\left[\Delta_i^{(2)} 1_{\{d_i>0\}} + 1 \cdot \mathbb{1}_{\{d_i=0\}}\right] \\ &= \mathbb{E}\left[\chi_i^{(2)} \mathbb{1}_{\{d_i>0\}}\right] - \mathbb{E}\left[\chi_i^{(1)} \mathbb{1}_{\{d_i>0\}}\right] + p_0.\end{aligned}$$

Using Lemmas 3.4 and 3.3 yields the result. □

Note that we have $\lim_{p\downarrow 0} p_0 = 1$, i.e. a vertex is isolated almost surely as $p \downarrow 0$, so that, for all i,

$$\lim_{p\downarrow 0} \mathbb{E}\left[\Delta_i^{(1)*}\right] = \lim_{p\downarrow 0} \mathbb{E}\left[\Delta_i^{(2)*}\right] = 1,$$

which is as expected since, by definition, $\Delta_i^{(k)*} = 1$ for i with $d_i = 0$, for all $k > 0$.

For the sparse case, in which $p = \frac{\lambda}{n-1}$ with $\lambda \in (0, \infty)$, we find the following results by substitution.

Corollary 3.7. *If $p = \frac{\lambda}{n-1}$ with $\lambda \in (0, \infty)$, then*

$$\mathbb{E}\left[\Delta_i^{(1)*}\right] \to 1 - \lambda e^{-\lambda}$$

as $n \to \infty$.

Corollary 3.8. *If $p = \frac{\lambda}{n-1}$ with $\lambda \in (0, \infty)$, then*

$$\mathbb{E}\left[\Delta_i^{(2)*}\right] \to \lambda^2(1 - e^{-\lambda}) + \lambda(1 + e^{-\lambda}) + 2e^{-\lambda} - 1$$

as $n \to \infty$.

We can also now calculate the difference between $\Delta_i^{(2)*}$ and $\Delta_i^{(1)*}$.

Corollary 3.9.

$$\begin{aligned}\mathbb{E}\left[\Delta_i^{(2)*} - \Delta_i^{(1)*}\right] &= n^2p^2(1 - p_0) + 5np^2p_0 - 4np^2 + np(1 + 2p_0) - 6p^2p_0 + 5p^2 - 2pp_0 - p \\ &\quad + 2p_0 - 2.\end{aligned}$$

Corollary 3.10. *If $p = \frac{\lambda}{n-1}$, then*

$$\mathbb{E}\left[\Delta_i^{(2)*} - \Delta_i^{(1)*}\right] \to \lambda^2(1 - e^{-\lambda}) + \lambda(1 + 2e^{-\lambda}) + 2e^{-\lambda} - 2$$

as $n \to \infty$.

Note that for $p = \frac{\lambda}{n-1}$,

$$\lim_{\lambda\downarrow 0} \mathbb{E}\left[\Delta_i^{(2)*} - \Delta_i^{(1)*}\right] = 0 \text{ and } \lim_{\lambda\to\infty} \mathbb{E}\left[\Delta_i^{(2)*} - \Delta_i^{(1)*}\right] = \infty,$$

as $n \to \infty$.

3.4 Chapter conclusion

In this chapter we discussed the Erdős–Rényi random graph for both the dense and the sparse case. For both cases the expectations of $\Delta_i^{(1)*}$ and $\Delta_i^{(2)*}$ were calculated.

4 Limit behavior

In this section we investigate the limit behavior of $\chi_i^{(k)}$ and $\Delta_i^{(k)*}$. What happens when k, which indicates the distance of the friends whose popularity we are interested in, or n, the amount of individuals in the network, tends to infinity?

4.1 Finite graph, large level

Let G be an undirected, simple and connected graph with n vertices. For a vertex i, we consider the behavior of

$$\chi_i^{(k)} = \frac{1}{d_i} \sum_{j=1}^{n} \left(A^k\right)_{ij} d_j$$

as $k \to \infty$.

Note that for $k \in \mathbb{N}$, the quantity $\left(A^k\right)_{ij}$ denotes the number of k-step paths from i to j. Hence,

$$\left(A^k\right)_{ij} \propto \pi_j \text{ as } k \to \infty,$$

where π is the stationary distribution of the simple random walk on the graph [6, Theorem 4]. Indeed, as $k \to \infty$, $\left(A^k\right)_{ij}$ is proportional to the probability of the random walk finding itself in vertex j, irrespective of the initial vertex i.

4.2 Large graph, finite level

By Theorems 3.5 and 3.6, for $p \in [0,1]$ and $n \in \mathbb{N}$, we have for $\mathrm{ER}_n(p)$ the results

$$\mathbb{E}\left[\Delta_i^{(1)*}\right] = 1 - (n-2)pp_0 - p$$

and

$$\mathbb{E}\left[\Delta_{[n]}^{(2)*}\right] = \mathbb{E}\left[\Delta_i^{(2)*}\right] = n^2p^2(1-p_0) + 5np^2p_0 - 4np^2 + np(1+p_0) - 6p^2p_0 + 5p^2 - 2p + 2p_0 - 1.$$

By Corollaries 3.7 and 3.8, for $p = \frac{\lambda}{n-1}$ with $\lambda \in (0,\infty)$, we have for $\mathrm{ER}_n(p)$ the results

$$\mathbb{E}\left[\Delta_i^{(1)*}\right] \to 1 - \lambda e^{-\lambda}$$

and

$$\mathbb{E}\left[\Delta_{[n]}^{(2)*}\right] = \mathbb{E}\left[\Delta_1^{(2)*}\right] \to \lambda^2(1-e^{-\lambda}) + \lambda(1+e^{-\lambda}) + 2e^{-\lambda} - 1$$

as $n \to \infty$.

We now seek to calculate $\operatorname{var} \Delta_i^{(1)*}$.

Lemma 4.1.

$$\mathbb{E}\left[\left(\Delta_i^{(1)*}\right)^2\right] = -4p_0p^2n^2 + 23p_0p^2n - 33p_0p^2 - 3p_0pn + 6p_0p - 3p^2n^2A + 3p^2n^2 + 18p^2nA - 20p^2n - 27p^2A + 31p^2 + 2pn - 5p + A + 1,$$

where

$$A = \sum_{k=1}^{n-1} \frac{p_k}{k}.$$

Proof. Compute

$$\begin{aligned}
\mathbb{E}\left[\left(\Delta_i^{(1)*}\right)^2\right] &= \mathbb{E}\left[\left(\Delta_i^{(1)}\mathbb{1}_{\{d_i>0\}} + \mathbb{1}_{\{d_i=0\}}\right)^2\right] \\
&= \mathbb{E}\left[\left(\Delta_i^{(1)}\mathbb{1}_{\{d_i>0\}}\right)^2 + \mathbb{1}_{\{d_i=0\}}\right] \\
&= \mathbb{E}\left[\left(\Delta_i^{(1)}\right)^2 \mathbb{1}_{\{d_i>0\}}\right] + p_0 \\
&= \mathbb{E}\left[\left(\frac{1}{d_i}\sum_{j=1}^n A_{ij}d_j - d_i\right)^2 \mathbb{1}_{\{d_i>0\}}\right] + p_0 \\
&= \mathbb{E}\left[\left(\frac{1}{d_i^2}\sum_{j=1}^n\sum_{l=1}^n A_{ij}A_{il}d_jd_l - 2\sum_{j=1}^n A_{ij}d_j + d_i^2\right)\mathbb{1}_{\{d_i>0\}}\right] + p_0 \qquad (7)
\end{aligned}$$

We compute the expectations of the terms separately.

For the first term in (7),

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{d_i^2}\sum_{j=1}^n\sum_{l=1}^n A_{ij}A_{il}d_jd_l\mathbb{1}_{\{d_i>0\}}\right] &= \mathbb{E}\left[\frac{1}{d_i^2}\sum_{j=1}^n\sum_{l=1}^n\sum_{h=1}^n\sum_{r=1}^n A_{ij}A_{il}A_{jh}A_{lr}\mathbb{1}_{\{d_i>0\}}\right] \\
&= \sum_{k=1}^{n-1}\frac{p_k}{k^2}\sum_{j=1}^n\sum_{l=1}^n\sum_{h=1}^n\sum_{r=1}^n \mathbb{E}\left[A_{ij}A_{il}A_{jh}A_{lr}|d_i=k\right].
\end{aligned}$$

In this sum we distinguish between cases as follows:

- $l = j, h \neq i, r = i$:

$$\begin{aligned}
\sum_{k=1}^{n-1}\frac{p_k}{k^2}\sum_{j=1}^n\sum_{\substack{h=1\\h\neq i}}^n \mathbb{E}\left[A_{ij}A_{jh}|d_i=k\right] &= \sum_{k=1}^{n-1}\frac{p_k}{k^2}\sum_{\substack{j=1\\j\neq i}}^n\sum_{\substack{h=1\\h\notin\{i,j\}}}^n \mathbb{E}\left[A_{ij}A_{jh}|d_i=k\right] \\
&= p(n-2)\sum_{k=1}^{n-1}\frac{p_k}{k^2}\sum_{\substack{j=1\\j\neq i}}^n \mathbb{E}\left[A_{ij}|d_i=k\right] \\
&= p(n-2)\sum_{k=1}^{n-1}\frac{p_k}{k}
\end{aligned}$$

- $l = j, h \neq i, r = h$:

$$\sum_{k=1}^{n-1}\frac{p_k}{k^2}\sum_{j=1}^n\sum_{\substack{h=1\\h\neq i}}^n \mathbb{E}\left[A_{ij}A_{jh}|d_i=k\right] = p(n-2)\sum_{k=1}^{n-1}\frac{p_k}{k}$$

- $l = j, h \neq i, r \notin \{i, h\}$:

$$\sum_{k=1}^{n-1} \frac{p_k}{k^2} \sum_{j=1}^{n} \sum_{\substack{h=1 \\ h\neq i}}^{n} \sum_{\substack{r=1 \\ r\notin\{i,h\}}}^{n} \mathbb{E}\left[A_{ij}A_{jh}A_{jr}|d_i = k\right] = \sum_{k=1}^{n-1} \frac{p_k}{k^2} \sum_{\substack{j=1 \\ j\neq i}}^{n} \sum_{\substack{h=1 \\ h\notin\{i,j\}}}^{n} \sum_{\substack{r=1 \\ r\notin\{i,j,h\}}}^{n} \mathbb{E}\left[A_{ij}A_{jh}A_{jr}|d_i = k\right]$$
$$= p^2(n-2)(n-3) \sum_{k=1}^{n-1} \frac{p_k}{k^2} \sum_{\substack{j=1 \\ j\neq i}}^{n} \mathbb{E}\left[A_{ij}|d_i = k\right]$$
$$= p^2(n-2)(n-3) \sum_{k=1}^{n-1} \frac{p_k}{k}$$

- $l = j, h = i, r = i$:

$$\sum_{k=1}^{n-1} \frac{p_k}{k^2} \sum_{j=1}^{n} \mathbb{E}\left[A_{ij}|d_i = k\right] = \sum_{k=1}^{n-1} \frac{p_k}{k}$$

- $l = j, h = i, r = h$:

$$\sum_{k=1}^{n-1} \frac{p_k}{k^2} \sum_{j=1}^{n} \mathbb{E}\left[A_{ij}|d_i = k\right] = \sum_{k=1}^{n-1} \frac{p_k}{k}$$

- $l = j, h = i, r \notin \{i, h\}$:

$$\sum_{k=1}^{n-1} \frac{p_k}{k^2} \sum_{j=1}^{n} \sum_{\substack{r=1 \\ r\neq i}}^{n} \mathbb{E}\left[A_{ij}A_{jr}|d_i = k\right] = \sum_{k=1}^{n-1} \frac{p_k}{k^2} \sum_{\substack{j=1 \\ j\neq i}}^{n} \sum_{\substack{r=1 \\ r\notin\{i,j\}}}^{n} \mathbb{E}\left[A_{ij}A_{jr}|d_i = k\right]$$
$$= p(n-2) \sum_{k=1}^{n-1} \frac{p_k}{k^2} \sum_{\substack{j=1 \\ j\neq i}}^{n} \mathbb{E}\left[A_{ij}|d_i = k\right]$$
$$= p(n-2) \sum_{k=1}^{n-1} \frac{p_k}{k}$$

- $l \neq j, h \notin \{i, l\}, r \notin \{i, j, h\}$:

$$\sum_{k=1}^{n-1} \frac{p_k}{k^2} \sum_{j=1}^{n} \sum_{\substack{l=1 \\ l\neq j}}^{n} \sum_{\substack{h=1 \\ h\notin\{i,l\}}}^{n} \sum_{\substack{r=1 \\ r\notin\{i,j,h\}}}^{n} \mathbb{E}\left[A_{ij}A_{il}A_{jh}A_{lr}|d_i = k\right]$$
$$= \sum_{k=1}^{n-1} \frac{p_k}{k^2} \sum_{\substack{j=1 \\ j\neq i}}^{n} \sum_{\substack{l=1 \\ l\notin\{i,j\}}}^{n} \sum_{\substack{h=1 \\ h\notin\{i,j,l\}}}^{n} \sum_{\substack{r=1 \\ r\notin\{i,j, \\ h,l\}}}^{n} \mathbb{E}\left[A_{ij}A_{il}A_{jh}A_{lr}|d_i = k\right]$$

$$= p^2(n-3)(n-4) \sum_{k=1}^{n-1} \frac{p_k}{k^2} \sum_{\substack{j=1 \\ j\neq i}}^{n} \sum_{\substack{l=1 \\ l\notin\{i,j\}}}^{n} \mathbb{E}\left[A_{ij}A_{il}|d_i = k\right],$$

with

$$\sum_{k=1}^{n-1}\frac{p_k}{k^2}\sum_{\substack{j=1\\j\neq i}}^{n}\sum_{\substack{l=1\\l\notin\{i,j\}}}^{n}\mathbb{E}\left[A_{ij}A_{il}|d_i=k\right]=\sum_{k=1}^{n-1}\frac{p_k}{k^2}\sum_{\substack{j=1\\j\neq i}}^{n}\mathbb{E}\left[A_{ij}(d_i-A_{ij})|d_i=k\right]$$

$$=\sum_{k=1}^{n-1}\frac{p_k}{k^2}\sum_{\substack{j=1\\j\neq i}}^{n}\mathbb{E}\left[kA_{ij}|d_i=k\right]-\sum_{k=1}^{n-1}\frac{p_k}{k^2}\sum_{\substack{j=1\\j\neq i}}^{n}\mathbb{E}\left[A_{ij}|d_i=k\right]$$

$$=\sum_{k=1}^{n-1}p_k-\sum_{k=1}^{n-1}\frac{p_k}{k}$$

$$=(1-p_0)-\sum_{k=1}^{n-1}\frac{p_k}{k},$$

so

$$\sum_{k=1}^{n-1}\frac{p_k}{k^2}\sum_{j=1}^{n}\sum_{\substack{l=1\\l\neq j}}^{n}\sum_{\substack{h=1\\h\notin\{i,l\}}}^{n}\sum_{\substack{r=1\\r\notin\{i,j,h\}}}^{n}\mathbb{E}\left[A_{ij}A_{il}A_{jh}A_{lr}|d_i=k\right]=p^2(n-3)(n-4)\left[(1-p_0)-\sum_{k=1}^{n-1}\frac{p_k}{k}\right]$$

- $l\neq j, h\notin\{i,l\}, r=i$:

$$\sum_{k=1}^{n-1}\frac{p_k}{k^2}\sum_{j=1}^{n}\sum_{\substack{l=1\\l\neq j}}^{n}\sum_{\substack{h=1\\h\notin\{i,l\}}}^{n}\mathbb{E}\left[A_{ij}A_{il}A_{jh}|d_i=k\right]=\sum_{k=1}^{n-1}\frac{p_k}{k^2}\sum_{\substack{j=1\\j\neq i}}^{n}\sum_{\substack{l=1\\l\notin\{i,j\}}}^{n}\sum_{\substack{h=1\\h\notin\{i,j,l\}}}^{n}\mathbb{E}\left[A_{ij}A_{il}A_{jh}|d_i=k\right]$$

$$=p(n-3)\sum_{k=1}^{n-1}\frac{p_k}{k^2}\sum_{\substack{j=1\\j\neq i}}^{n}\sum_{\substack{l=1\\l\notin\{i,j\}}}^{n}\mathbb{E}\left[A_{ij}A_{il}|d_i=k\right]$$

$$=p(n-3)\left[(1-p_0)-\sum_{k=1}^{n-1}\frac{p_k}{k}\right]$$

- $l\neq j, h\notin\{i,l\}, r=j$:

$$\sum_{k=1}^{n-1}\frac{p_k}{k^2}\sum_{j=1}^{n}\sum_{\substack{l=1\\l\neq j}}^{n}\sum_{\substack{h=1\\h\notin\{i,l\}}}^{n}\mathbb{E}\left[A_{ij}A_{il}A_{jh}A_{lj}|d_i=k\right]=\sum_{k=1}^{n-1}\frac{p_k}{k^2}\sum_{\substack{j=1\\j\neq i}}^{n}\sum_{\substack{l=1\\l\notin\{i,j\}}}^{n}\sum_{\substack{h=1\\h\notin\{i,j,l\}}}^{n}\mathbb{E}\left[A_{ij}A_{il}A_{jh}A_{lj}|d_i=k\right]$$

$$=p^2(n-2)(n-3)\sum_{k=1}^{n-1}\frac{p_k}{k^2}\sum_{\substack{j=1\\j\neq i}}^{n}\sum_{\substack{l=1\\l\notin\{i,j\}}}^{n}\mathbb{E}\left[A_{ij}A_{il}|d_i=k\right]$$

$$=p^2(n-2)(n-3)\left[(1-p_0)-\sum_{k=1}^{n-1}\frac{p_k}{k}\right]$$

- $l \neq j, h \notin \{i,l\}, r = h$:

$$\sum_{k=1}^{n-1} \frac{p_k}{k^2} \sum_{j=1}^{n} \sum_{\substack{l=1 \\ l\neq j}}^{n} \sum_{\substack{h=1 \\ h\notin\{i,l\}}}^{n} \mathbb{E}\left[A_{ij}A_{il}A_{jh}A_{lh}|d_i = k\right] = \sum_{k=1}^{n-1} \frac{p_k}{k^2} \sum_{\substack{j=1 \\ j\neq i}}^{n} \sum_{\substack{l=1 \\ l\notin\{i,j\}}}^{n} \sum_{\substack{h=1 \\ h\notin\{i,j,l\}}}^{n} \mathbb{E}\left[A_{ij}A_{il}A_{jh}A_{lh}|d_i = k\right]$$

$$= p^2(n-3)^2 \sum_{k=1}^{n-1} \frac{p_k}{k^2} \sum_{\substack{j=1 \\ j\neq i}}^{n} \sum_{\substack{l=1 \\ l\notin\{i,j\}}}^{n} \mathbb{E}\left[A_{ij}A_{il}|d_i = k\right]$$

$$= p^2(n-3)^2 \left[(1-p_0) - \sum_{k=1}^{n-1} \frac{p_k}{k}\right]$$

- $l \neq j, h = l, r \notin \{i,j\}$:

$$\sum_{k=1}^{n-1} \frac{p_k}{k^2} \sum_{j=1}^{n} \sum_{\substack{l=1 \\ l\neq j}}^{n} \sum_{\substack{r=1 \\ r\notin\{i,j\}}}^{n} \mathbb{E}\left[A_{ij}A_{il}A_{jl}A_{lr}|d_i = k\right] = \sum_{k=1}^{n-1} \frac{p_k}{k^2} \sum_{\substack{j=1 \\ j\neq i}}^{n} \sum_{\substack{l=1 \\ l\notin\{i,j\}}}^{n} \sum_{\substack{r=1 \\ r\notin\{i,j,l\}}}^{n} \mathbb{E}\left[A_{ij}A_{il}A_{jl}A_{lr}|d_i = k\right]$$

$$= p^2(n-2)(n-3) \sum_{k=1}^{n-1} \frac{p_k}{k^2} \sum_{\substack{j=1 \\ j\neq i}}^{n} \sum_{\substack{l=1 \\ l\notin\{i,j\}}}^{n} \mathbb{E}\left[A_{ij}A_{il}|d_i = k\right]$$

$$= p^2(n-2)(n-3) \left[(1-p_0) - \sum_{k=1}^{n-1} \frac{p_k}{k}\right]$$

- $l \neq j, h = l, r = i$:

$$\sum_{k=1}^{n-1} \frac{p_k}{k^2} \sum_{j=1}^{n} \sum_{\substack{l=1 \\ l\neq j}}^{n} \mathbb{E}\left[A_{ij}A_{il}A_{jl}|d_i = k\right] = \sum_{k=1}^{n-1} \frac{p_k}{k^2} \sum_{\substack{j=1 \\ j\neq i}}^{n} \sum_{\substack{l=1 \\ l\notin\{i,j\}}}^{n} \mathbb{E}\left[A_{ij}A_{il}A_{jl}|d_i = k\right]$$

$$= p(n-2) \sum_{k=1}^{n-1} \frac{p_k}{k^2} \sum_{\substack{j=1 \\ j\neq i}}^{n} \sum_{\substack{l=1 \\ l\notin\{i,j\}}}^{n} \mathbb{E}\left[A_{ij}A_{il}|d_i = k\right]$$

$$= p(n-2) \left[(1-p_0) - \sum_{k=1}^{n-1} \frac{p_k}{k}\right]$$

- $l \neq j, h = l, r = j$:

$$\sum_{k=1}^{n-1} \frac{p_k}{k^2} \sum_{j=1}^{n} \sum_{\substack{l=1 \\ l\neq j}}^{n} \mathbb{E}\left[A_{ij}A_{il}A_{jl}|d_i = k\right] = \sum_{k=1}^{n-1} \frac{p_k}{k^2} \sum_{\substack{j=1 \\ j\neq i}}^{n} \sum_{\substack{l=1 \\ l\notin\{i,j\}}}^{n} \mathbb{E}\left[A_{ij}A_{il}A_{jl}|d_i = k\right]$$

$$= p \sum_{k=1}^{n-1} \frac{p_k}{k^2} \sum_{\substack{j=1 \\ j\neq i}}^{n} \sum_{\substack{l=1 \\ l\notin\{i,j\}}}^{n} \mathbb{E}\left[A_{ij}A_{il}|d_i = k\right]$$

$$= p \left[(1-p_0) - \sum_{k=1}^{n-1} \frac{p_k}{k}\right]$$

- $l \neq j, h = i, r \notin \{i, j\}$:

$$\sum_{k=1}^{n-1} \frac{p_k}{k^2} \sum_{j=1}^{n} \sum_{\substack{l=1 \\ l\neq j}}^{n} \sum_{\substack{r=1 \\ r\notin\{i,j\}}}^{n} \mathbb{E}\left[A_{ij}A_{il}A_{lr}|d_i = k\right] = \sum_{k=1}^{n-1} \frac{p_k}{k^2} \sum_{\substack{j=1 \\ j\neq i}}^{n} \sum_{\substack{l=1 \\ l\notin\{i,j\}}}^{n} \sum_{\substack{r=1 \\ r\notin\{i,j,l\}}}^{n} \mathbb{E}\left[A_{ij}A_{il}A_{lr}|d_i = k\right]$$
$$= p(n-3) \sum_{k=1}^{n-1} \frac{p_k}{k^2} \sum_{\substack{j=1 \\ j\neq i}}^{n} \sum_{\substack{l=1 \\ l\notin\{i,j\}}}^{n} \mathbb{E}\left[A_{ij}A_{il}|d_i = k\right]$$
$$= p(n-3) \left[(1-p_0) - \sum_{k=1}^{n-1} \frac{p_k}{k}\right]$$

- $l \neq j, h = i, r = i$:

$$\sum_{k=1}^{n-1} \frac{p_k}{k^2} \sum_{j=1}^{n} \sum_{\substack{l=1 \\ l\neq j}}^{n} \mathbb{E}\left[A_{ij}A_{il}|d_i = k\right] = \sum_{k=1}^{n-1} \frac{p_k}{k^2} \sum_{\substack{j=1 \\ j\neq i}}^{n} \sum_{\substack{l=1 \\ l\notin\{i,j\}}}^{n} \mathbb{E}\left[A_{ij}A_{il}|d_i = k\right]$$
$$= (1-p_0) - \sum_{k=1}^{n-1} \frac{p_k}{k}$$

- $l \neq j, h = i, r = j$:

$$\sum_{k=1}^{n-1} \frac{p_k}{k^2} \sum_{j=1}^{n} \sum_{\substack{l=1 \\ l\neq j}}^{n} \mathbb{E}\left[A_{ij}A_{il}A_{lj}|d_i = k\right] = \sum_{k=1}^{n-1} \frac{p_k}{k^2} \sum_{j=1}^{n} \sum_{\substack{l=1 \\ l\neq j}}^{n} \mathbb{E}\left[A_{ij}A_{il}A_{jl}|d_i = k\right]$$
$$= p\left[(1-p_0) - \sum_{k=1}^{n-1} \frac{p_k}{k}\right]$$

For the second term in (7),

$$\mathbb{E}\left[-2\sum_{j=1}^{n} A_{ij}d_j \mathbb{1}_{\{d_i>0\}}\right] = -2\mathbb{E}\left[\sum_{j=1}^{n}\sum_{l=1}^{n} A_{ij}A_{jl}\mathbb{1}_{\{d_i>0\}}\right]$$
$$= -2\sum_{k=1}^{n-1} p_k \sum_{j=1}^{n}\sum_{l=1}^{n} \mathbb{E}\left[A_{ij}A_{jl}|d_i = k\right].$$

In this sum we distinguish the cases $l = i$ and $l \neq i$ as follows:

- $l \neq i$:

$$
\begin{aligned}
-2\sum_{k=1}^{n-1} p_k \sum_{j=1}^{n} \sum_{\substack{l=1\\ l\neq i}}^{n} \mathbb{E}\left[A_{ij}A_{jl} | d_i = k\right] &= -2\sum_{k=1}^{n-1} p_k \sum_{\substack{j=1\\ j\neq i}}^{n} \sum_{\substack{l=1\\ l\notin\{i,j\}}}^{n} \mathbb{E}\left[A_{ij}A_{jl} | d_i = k\right] \\
&= -2p(n-2)\sum_{k=1}^{n-1} p_k \sum_{\substack{j=1\\ j\neq i}}^{n} \mathbb{E}\left[A_{ij} | d_i = k\right] \\
&= -2p(n-2)\sum_{k=1}^{n-1} p_k k \\
&= -2p(n-2)\sum_{k=0}^{n-1} p_k k \\
&= -2p^2(n-1)(n-2)
\end{aligned}
$$

- $l = i$:

$$
\begin{aligned}
-2\sum_{k=1}^{n-1} p_k \sum_{j=1}^{n} \mathbb{E}\left[A_{ij} | d_i = k\right] &= -2\sum_{k=1}^{n-1} p_k \sum_{\substack{j=1\\ j\neq i}}^{n} \mathbb{E}\left[A_{ij} | d_i = k\right] \\
&= -2\sum_{k=1}^{n-1} p_k k \\
&= -2\sum_{k=0}^{n-1} p_k k \\
&= -2p(n-1)
\end{aligned}
$$

For the third term in (7),

$$
\mathbb{E}\left[d_i^2 \mathbb{1}_{\{d_i>0\}}\right] = \mathbb{E}\left[d_i^2\right] = (n-1)p(1-p) + (n-1)^2p^2.
$$

Addition of the terms yields the expectation. □

Now the variance of $\Delta_i^{(1)*}$ can be computed.

Theorem 4.2.

$$
\begin{aligned}
\operatorname{var}\Delta_i^{(1)*} = {}& 4p_0^2p^2n - p_0^2p^2n^2 - 4p_0^2p^2 - 4p_0p^2n^2 + 21p_0p^2n - 29p_0p^2 + 2p_0p - p_0pn - 3p^2n^2A \\
&+ 3p^2n^2 + 18p^2nA - 20p^2n - 27p^2A + 30p^2 + 2pn - 3p + A,
\end{aligned}
$$

where

$$
A = \sum_{k=1}^{n-1} \frac{p_k}{k}.
$$

Proof. This follows from the fact that

$$
\operatorname{var}\Delta_i^{(1)*} = \mathbb{E}\left[\left(\Delta_i^{(1)*}\right)^2\right] - \mathbb{E}\left[\Delta_i^{(1)*}\right]^2.
$$

□

Corollary 4.3. *If* $p = \frac{\lambda}{n-1}$, *then*

$$\operatorname{var} \Delta_i^{(1)*} \to \lambda^2 \left(e^{-2\lambda} - 4e^{-\lambda} - 3 \sum_{k=1}^{\infty} \frac{\lambda^k e^{-\lambda}}{k \cdot k!} + 3 \right) + \lambda \left(2 - e^{-\lambda}\right) + \sum_{k=1}^{\infty} \frac{\lambda^k e^{-\lambda}}{k \cdot k!}$$

as $n \to \infty$.

Note that for $p = \frac{\lambda}{n-1}$,

$$\lim_{\lambda \downarrow 0} \operatorname{var} \Delta_i^{(1)*} = 0 \text{ and } \lim_{\lambda \to \infty} \operatorname{var} \Delta_i^{(1)*} = \infty,$$

as $n \to \infty$.

4.3 Chapter conclusion

This chapter dealth with the limit behavior of $\chi_i^{(k)}$ and $\Delta_i^{(k)*}$. We considered what happens in a finite graph with a large level of friendship, and in a large graph with a finite level of friendship. The variance of $\Delta_i^{(1)*}$ was also calculated, both for the dense and the sparse case of the Erdős–Rényi random graph.

5 Conclusion

In Chapter 1, some preliminaries along with the statement and proof of the classical Friendship Paradox were given. The phenomenon was also illustrated by an example.

Chapter 2 dealt with the extension of the Friendship Paradox to higher levels of friendship. Relevant quantities were introduced and it was proved that the Friendship Paradox holds for higher levels as well. Additionally, we considered graphs with isolated vertices by a modification of the $\Delta_i^{(k)}$. The same results turned out to hold still.

The Erdős–Rényi random graph model was introduced in Chapter 3, along with calculations of the expectations of $\Delta_i^{(1)*}$ and $\Delta_i^{(2)*}$ in such random graphs. These expectations were calculated both for the dense case and the sparse case.

In Chapter 4 we looked at the limit behavior of $\chi_i^{(k)}$ and $\Delta_i^{(k)*}$. We also calculated the variance of $\Delta_i^{(1)*}$.

Looking forward, we propose two avenues for further research:

- The only random graph model we investigated was the Erdős–Rényi random graph model in Chapter 3. It would be interesting to look at other random graph models, such as the inhomogeneous Erdős–Rényi random graph model, in which the connection probabilities is dependent on the pairs of vertices, or the Configuration Model (see [3, §3.1]).
- Which results do we get if k is dependent on n?

References

[1] Feld, S. L. (1991). Why Your Friends Have More Friends Than You Do. *American Journal of Sociology, 96*(6), 1464–1477. https://www.jstor.org/stable/2781907

[2] Cantwell, G. E., Kirkley, A., & Newman, M. (2021). The friendship paradox in real and model networks. *Journal of Complex Networks, 9*(2). https://doi.org/10.1093/comnet/cnab011

[3] MacDonald, P. (2022). *The Friendship Paradox* [Bachelor thesis, Leiden University].

[4] Kojaku, S., Hébert-Dufresne, L., Mones, E., Lehmann, S., & Ahn, Y. Y. (2021). The effectiveness of backward contact tracing in networks. *Nature physics, 17*, 652–658. https://doi.org/10.1038/s41567-021-01187-2

[5] *"What Do Your Friends Think?": Efficient Polling Methods for Networks Using Friendship Paradox*. (2021, March 1). IEEE Journals & Magazine — IEEE Xplore. https://ieeexplore.ieee.org/document/8832260

[6] Nikolov, A. (2019). *Week 5: Random Walks and Markov Chains* [Lecture notes]. http://www.cs.toronto.edu/~anikolov/CSC473W20/Lectures/MarkovChains.pdf