

**Team Members-** Akash, Ariel, Prabhpahul, Sundar & Wesley

## **Summary of the results of the preliminary analysis and data description**

**Our client is an author who writes fiction novels and wants to target the same age group audience as Harry Potter novels. The author wants to do some longitudinal analysis on the data sets so that he can get some inspiration before starting his own novel. The main purpose of this analysis is to do some research for this author(our client) for him to get a good understanding of the complex relationship dynamics between different characters of this novel. This would help the author to get ideas and derive some inspiration from one of the most famous Novel series in the world and he can take inspiration from this knowledge while writing his own book.**

We studied relationships between various characters of the Novel series Harry Potter. We studied various network characteristics like homophily, support groups among different characters and more. After exploring multiple topics we landed on this amazing much loved Harry Potter series with data and reference scripts found online on <http://www.stats.ox.ac.uk> and on github. This theme also helped us discover peer to peer network statistics and support groups development among adolescents. Among others, we studied structural characteristics like reciprocity and transitivity, we discovered that students with similar personality traits do significantly support each other and also since the school is a boarding school and is in an isolated area, students of all ages and gender mostly spent their spare time together.

Since this is a fictional story and all the data was limited to what the author outlined the world of Harry Potter and his close friends, the interactions between side characters are less known and consequently there are limitations on the dataset from this perspective. We have mentioned below the in-depth analysis of our study with questions and answers and visualizations.

## **The in-depth questions we are going to answer**

We formulated four questions to be answered with RSiena as they deal with (1) the way that fixed covariates of the characters affect the likelihood of forming support ties from the perspective of both alter and ego and (2) a longitudinal study of the variation in the likelihood that characters form ties across time. The questions to answer are:

1. Is there reciprocity in support among characters affected by gender? That is to say, if A and B are the same gender, are they more likely to show mutual support?
2. Are support ties more likely over time than expected by random chance?
3. Are women more likely to receive support from others in the school regardless of the gender or any other fixed covariate of the supporter?
4. Is there a tendency for students to demonstrate support to other students in their respective houses or do they equally support students from other houses?

Our client is an author who writes fiction novels and wants to target the same age group audience as Harry Potter novels. The author wants to do some longitudinal analysis on the data sets so that he can get some inspiration before starting his own novel. The main purpose of this analysis is to do some research for this author(our client) for him to get a good understanding of the complex relationship dynamics between different characters of this novel. This would help the author to get ideas and derive some inspiration from one of the most famous Novel series in the world and he can take inspiration from this knowledge while writing his own book. In the Harry Potter universe, personal relationships among characters, and thus support networks, are challenged by extreme circumstances only observable in a mystic novel. For example, when Voldemort (villain) becomes a threat to the magical world, characters must align against or in-favour of Voldemort thus delimiting sides. Situations like that arise constantly throughout the books and force characters to either form friendships and alliances or become enemies. For that reason, our client believes there is value in answering such questions. Our client wants us to analyse support among characters as differences between houses become pronounced and trust ability is put in jeopardy when Voldemort comes back. Our client informed us that Harry Potter fans often like to discuss which house they would belong to if they were to go to Hogwarts and they want to share this content with them as it will bolster discussions and strengthen their sense of 'belonging' to a particular house.

## Analysis

In order to address the findings, we used a small batch of both default and included effects. Among the included effects were *sameXRecip* for same-covariate (gender) reciprocity, *altX* for evaluation of the gender alter, and *sameX* for homophily based on the individual's house. We also utilized default effects—rates of constant friendship rate (period 1 & 2) and outdegree density—in the model in order to analyze and make conclusions to the questions.

For the first question, we aim to test if, in the case that they are the same gender, are A and B more likely to show mutual support? *SameXRecip* parameter is a measure of reciprocity based on the covariate value, which in this case is the value for gender. In the dataset, gender is denoted between 1 and 2, 1 being male and 2 being female. *SameXRecip* tracks reciprocity based on the same covariate (gender) value. The line for the effect is indicated below:

```
# First question  
myeff <- includeEffects(myeff, sameXRecip, interaction1 = "gender")
```

In the model, the *sameXRecip* parameter operates in that a positive estimation likely points to the existence of reciprocity based on the covariate value. If the model results in a negative value, then it signals the lack of reciprocity. However, the analysis of the effect ultimately relies on both the estimation and its statistical significance. If, for instance, the estimation provided by the parameter is positive and statistically significant, we answer our questions by concluding that there *is* a tendency for reciprocity between persons A and B, if they both share the same gender covariate value. The same analysis works conversely, and so on.

The third question asks if women are more likely to receive support from others, regardless of any covariate effects including gender. *AltX* is the parameter in RSiena that is used to measure covariate-related popularity. We use this parameter because we aim to find if there is a covariate-related popularity with regards to females. The line for this effect is seen below:

```
# Third question  
myeff <- includeEffects(myeff, altX, interaction1 = "gender")
```

Like the first question, this effect is based on the attribute for gender. As noted before, the dataset notes males as 1 and females as 2. Therefore in the model, when the *altX* parameter results in an estimation that is positive, we interpret that estimation as a popularity effect with regards to males, while negative estimation notes a popularity effect with regards to women.

## Findings

Running RSiena with the specified effects yielded the following parameter estimates:

Estimates, standard errors and convergence t-ratios

	Estimate	Standard Error	Convergence t-ratio
1. rate constant friendship rate (period 1)	7.1463	( 1.7206 )	-0.0247
2. rate constant friendship rate (period 2)	2.5263	( 0.5238 )	0.0391
3. eval outdegree (density)	-6.0546	( 0.5417 )	-0.0198
4. eval reciprocity	4.3688	( 0.4892 )	0.0116
5. eval gender alter	0.2111	( 0.2569 )	0.0614
6. eval same gender x reciprocity	-0.1884	( 0.4354 )	-0.0009
7. eval same house	2.9431	( 0.4880 )	-0.0098

Overall maximum convergence ratio: 0.1090

Total of 2355 iteration steps.

The parameter reciprocity converged as its convergence t-ratio is below 0.1, the conventional threshold. Its value is 4.3688 and the standard error 0.5417, which implies that it converged to a statistically significant value since estimate/SE > 1.96. We can interpret from this that if a character A supports character B then the odds of B forming a tie with A are  $\exp(4.368) = 78.9$  times higher than not forming a tie. This uncovers that it is extremely likely that a character reciprocates support, in other words: if a character shows support to someone then they can expect to receive back support from them. In the case of gender-dependent reciprocity, the parameter did not reach statistical significance which shows that reciprocated support does not depend on gender.

The parameter gender later converged to a reliable value since the t-ratio is less than 0.1. Its value was 0.211 and the standard error 0.2569 which implies that it was not statistically significant as estimate/SE < 1.96 in this case. Interestingly, women were equally as likely as men to receive support from others.

The parameter same house converged to a reliable value as its convergence r-ratio is below 0.1. Its value was 2.9431 and the standard error 0.488 which implies that it converged to a

statistically significant value since  $\text{estimate}/\text{SE} > 1.96$ . This implies that people are more  $\exp(2.9431)=18.97$  times more likely to show support to others in their own house than not.

The parameter outdegree converged to a reliable value as its convergence r-ratio is below 0.1. Its value was -6.0546 and the standard error 0.5417 which implies that it converged to a statistically significant value since  $\text{estimate}/\text{SE} > 1.96$ . This shows that in the Harry Potter saga, characters are less  $\exp(-6.054) = 0.002$  times less likely to show support to others than by chance. This uncovers that characters are generally not very supportive of each other.

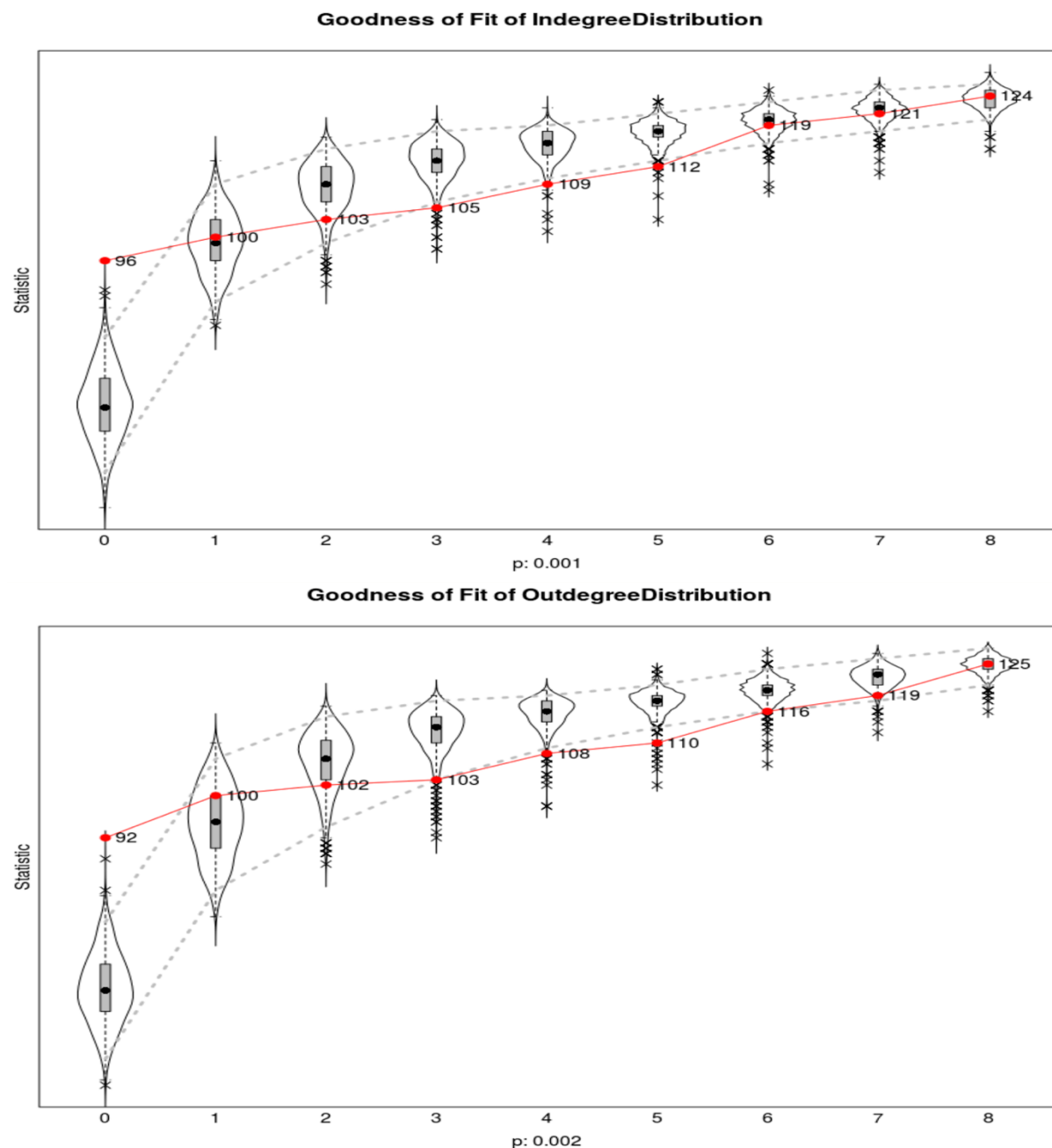


Figure 1: Goodness of fit graphs

## Implications

Our client is an author who writes fiction novels and wants to target the same age group audience as Harry Potter novels. The author wants to do some longitudinal analysis on the data sets so that he can get some inspiration before starting his own novel. The main purpose of this analysis is to do some research for this author(our client) for him to get a good understanding of the complex relationship dynamics between different characters of this novel. This would help the author to get ideas and derive some inspiration from one of the most famous Novel series in the world and he can take inspiration from this knowledge while writing his own book.

We answered a few questions for our client. The questions were -

1. Is there reciprocity in support among characters affected by gender? That is to say, if A and B are the same gender, are they more likely to show mutual support?
2. Are support ties more likely over time than expected by random chance?
3. Are women more likely to receive support from others in the school regardless of the gender or any other fixed covariate of the supporter?
4. Is there a tendency for students to demonstrate support to other students in their respective houses or do they equally support students from other houses?

Now with the help of these answers our client can get an in-depth understanding about the relationship between different characters in the book. Now since this is one of the best-selling novel series in the world, such knowledge could provide inspiration to future writers like our client.

Our recommendation for our client would be to properly analyse the questions that we have answered. For example we came to a conclusion that people are 18.97 times more likely to show support to others in their own house than not. Our client can take creative liberty to interpret this and apply it in some other situations. Similarly, there is a high level of reciprocity in showing support between characters. Now there is no hard and fast rule for our client to have exactly the same levels of reciprocity in his novels but this kind of analysis prepares an author and expands his horizon so that he can be more creative.

### **Recommendation to our Client**

Since it is a creative space, there are not hard and fast rules as to what makes a series interesting and successful. However, this Harry Potter series is one of the most famous book series in the world and had die-hard fan following. Taking some indirect and vague inspirations from this could prove quite beneficial.

We would recommend the author(our client) to try to incorporate similar patterns into his writing. Some of the patterns that the client can incorporate into the writing can be-

1. If a character shows support to someone then they can expect to receive back support from them.
2. Women were equally as likely as men to receive support from others.
3. People are more likely to show support to others in their own house than not.
4. Characters are generally not very supportive of each other.

## Reflection

As cited in the implications section the results drawn from our analysis with Siena and the EstAlgorithm, enabled us to address the questions and gather some insights. These have been tabulated and mentioned in detail in the Findings.

In regard to the metrics and scores obtained, we can note that reciprocity is a theme prevalent in the books, and this is rather intuitive to make a narrative with engaging characters and form ties and explore relations beyond the key nodes(main characters). In regard to the questions as a whole, we were able to address them and derive the relationship dynamics, and also note that reciprocity always does not pertain to the same gender(homophily) and extended beyond the groups as well.

An aspect to note from the visualizations is that across the three books, the ties formed or shown vary a lot. It can be attributed to the central characters and their interactions, whereas for Book 2 (Chamber of Secrets) it delved into arcs beyond the core group and formed a cluster and network beyond as illustrated. Moving forward to 3, we can note that the Azkaban arc mostly was insular to Harry and his peer group, mainly to the events that are covered in the book. The goodness of fit results can be noted to be good as well.

From our analysis we interpret that if a character A supports character B then the odds of B forming a tie with A is 78.9 times higher than not forming a tie. This uncovers that it is extremely likely that a character reciprocates support, in other words: if a character shows support to someone then they can expect to receive back support from them. We found out in our analysis that reciprocated support does not depend on gender. This shows us that even though the characters in the novel appear self centered, they have a high sense of reciprocity. If the author wants to take inspiration from this storyline, the characters in the story should also follow the same pattern.

Based on our finding we found out that women were equally as likely as men to receive support from others. After our analysis the widespread idea that there would be some difference in receiving support based on gender was dismissed. This finding is quite interesting since we intuitively think that there would be some degree of difference in the way different gender received support but that is not the case. The author should take this into account when forming up with his ideas.

After we conducted analysis on the dataset, we found out that people are more 18.97 times more likely to show support to others in their own house than not. The client can take creative liberty to interpret this and apply it in some other situations where this kind of storyline exists. For example, in a particular storyline, if we have a group of character in a scene, the characters should be supportive to each other within the group.

We found out that in the Harry Potter saga, characters are 0.002 times less likely to show support to others than by chance. This uncovers that characters are generally not very supportive of each other. This finding truly shows us the complex nature of the character in the Harry Potter series. This complexity is what made them so unique and interesting. In our previous finding we saw that the characters had a high sense of reciprocity, however in this finding we could see that they are quite self centered and isolated in some sense. This kind of complexity is quite intriguing and the author can take some inspiration from it.

**Visualizations on next page**



# Visualizations

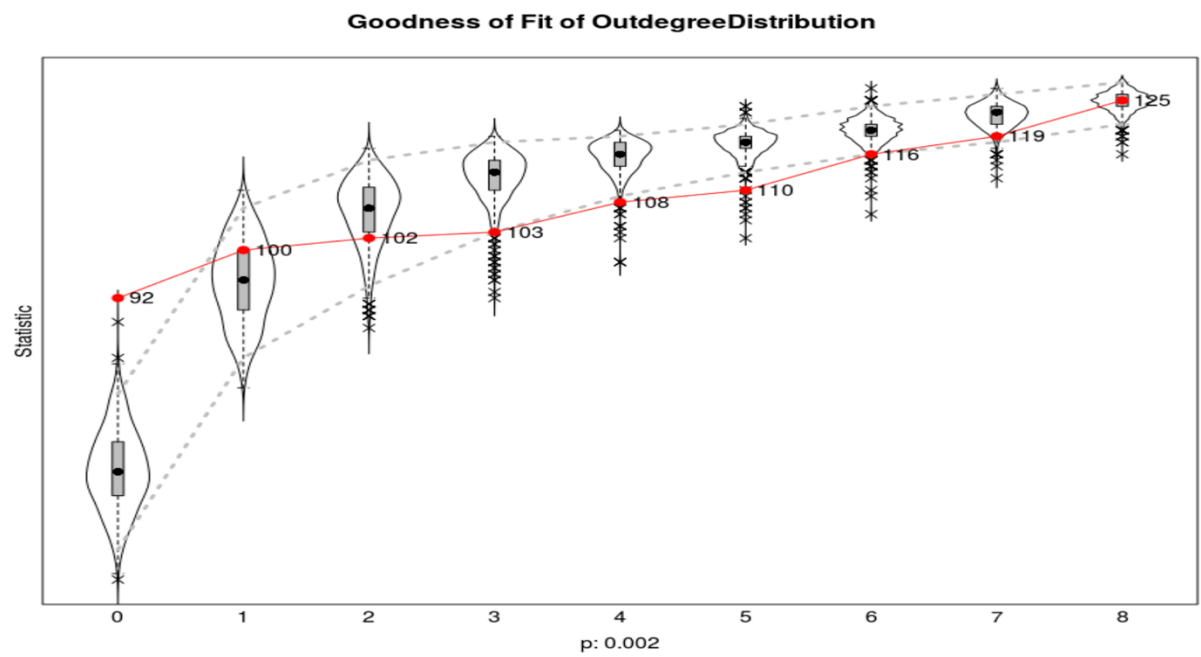
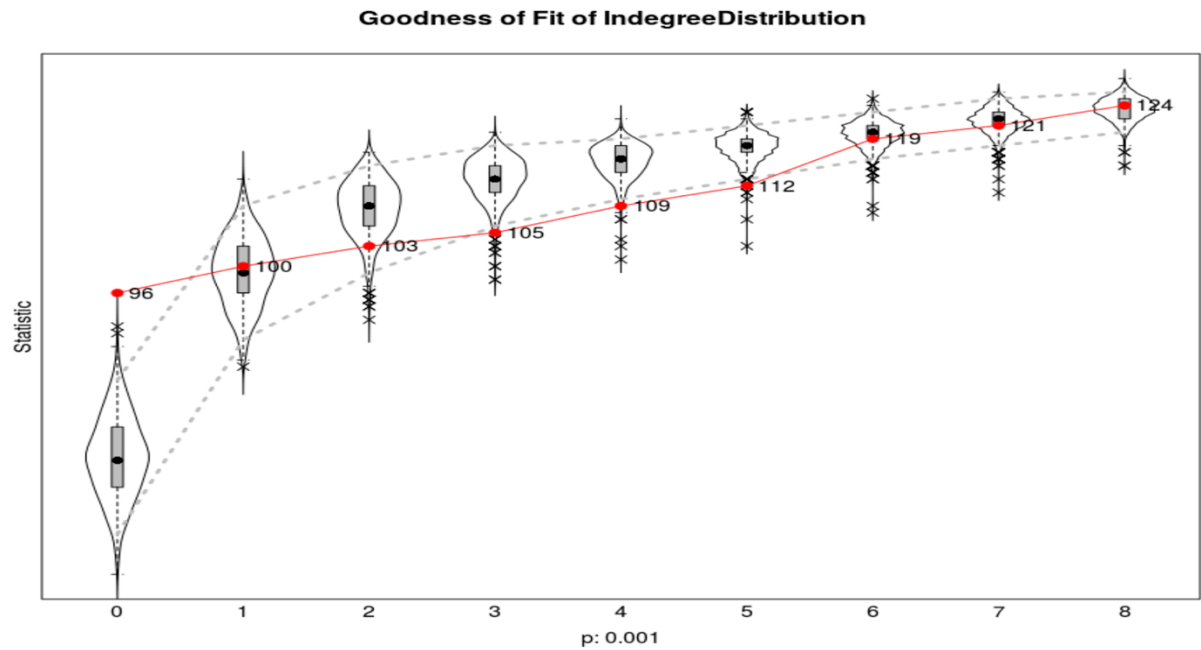


Figure 1: Goodness of fit plots for in-degree & out-degree distributions

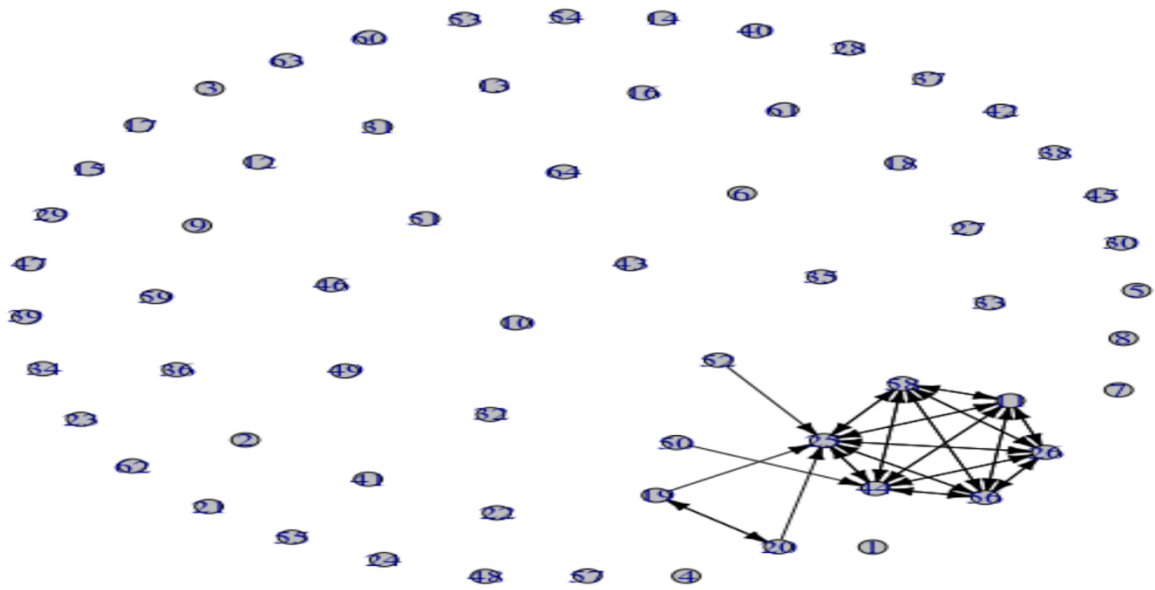


Figure 2: Support network from book 1

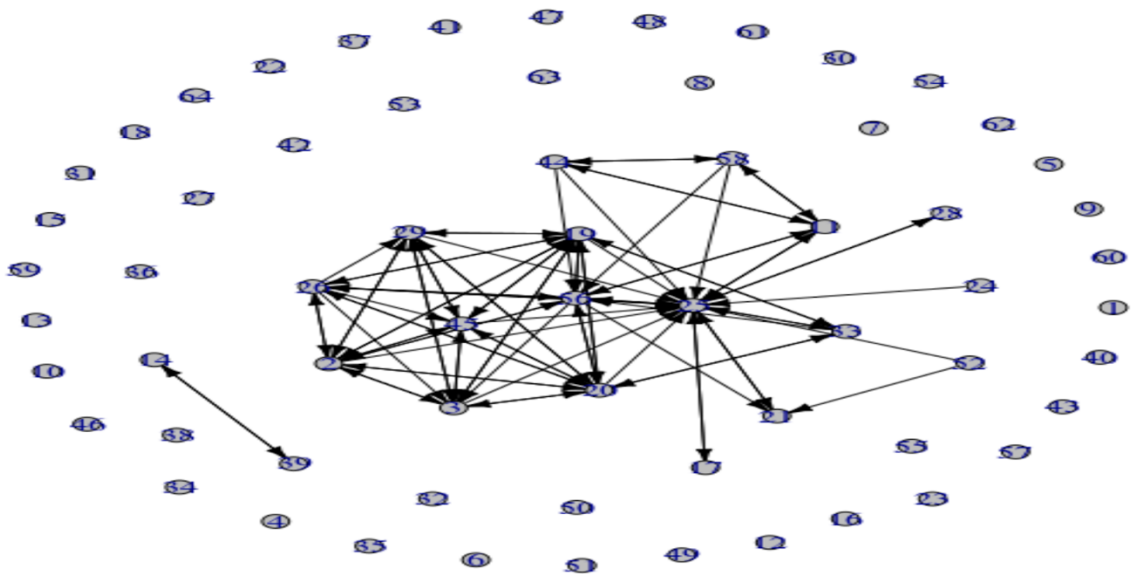


Figure 3: Support network from book 2

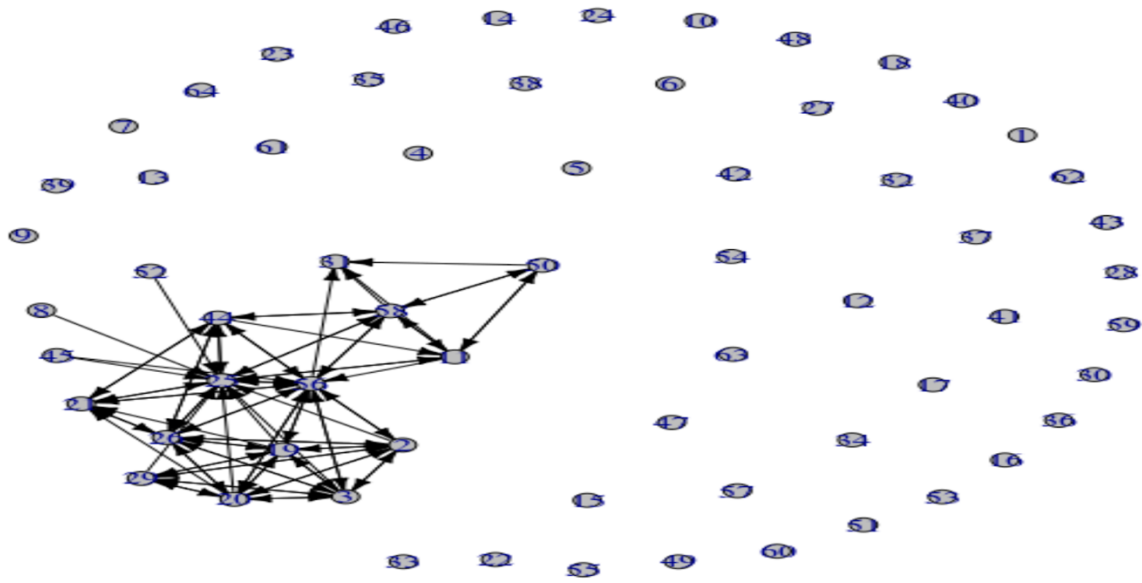


Figure 4: Support network from book 3