Proceedings of CONF-CIAP 2026 Symposium: International Conference on Atomic Magnetometer and Applications
DOI: 10.54254/2753-8818/2026.HZ27624

Social Networks Count Data: Negative Binomial Distributions and Extensions Versus Poisson and Bernoulli Models

Zhanyu Wan

Wuhan Britain-China School, Wuhan, China
wzyraymond@outlook.com

Abstract. With the development of the modern Internet, the use of social networks has increased significantly, which has promoted the research on social network analysis models in modern times. However, the modern academic community lacks a systematic integration and analysis of the selection of social network models, making it difficult for researchers and practitioners to choose a suitable model according to the actual situation of their research. This study aims to explain the advantages and disadvantages of four social network models (Bernoulli, Poisson, negative binomial, and zero-inflated) and systematically illustrate which model is suitable for which social network analysis through mathematical calculations and tabular organization. This paper concludes that in binary analysis and traditional analysis, the Bernoulli distribution and Poisson distribution are suitable for explaining such simple social networks. However, as the complexity of social networks increases, when complex conditions such as the proliferation of zero values and overdispersion are added, the shortcomings of simple models are revealed, which the more complex negative binomial model and zero-inflated model can well calculate and analyze social networks. This study provides a systematic model selection strategy for future researchers, but it lacks the analysis of dynamic multi-model social networks and mainly focuses on models of static networks. Therefore, future researchers can focus on analyzing the application of more complex models in dynamic networks and their advantages and disadvantages.

Keywords: Social network, Bernoulli distribution, Poisson distribution, negative binomial distribution,zero-inflated negative binomial distribution.

1.Introduction

In recent years, due to the explosive growth of social networks, the data among people's relationships has shown a flood of zero values. Many people do not have actual interactions, and there is overdispersion, with the variance being much larger than the expectation, and the generation mechanism is complex. The limitations of traditional models have gradually become prominent. This paper looks at how social networks are getting more complicated and finds better ways to analyze them than the old methods. It not only complements the development context of statistical theory but also provides an underlying basis for model selection in real-life scenarios.

© 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2026.HZ27624

Due to the increasing complexity of social network analysis, people cannot fully understand when to use the Bernoulli distribution, when to use the Poisson distribution, and when to use the negative binomial distribution. People often struggle to really grasp why the negative binomial distribution is used to tackle certain problems. When talking about social networks and trying to understand them, the line between what's actually being used and the math behind it often gets a bit blurry. There is a lack of clarity regarding which model to use, and the importance of model expansion in complex networks has not been deeply understood. This has created some research gaps. The connection between model features, mathematical logic, and the demands of real-world social networks has not been explicitly proposed and explained in previous papers. Researchers lack suggestions on how to choose the appropriate model, which hinders the accurate analysis of complex network data.

This paper challenges the usual way of looking at models on their own. By introducing the advantages and disadvantages of different distributions, it explains how the underlying mathematical logic solves the complexity of social networks, compares the contributions of various distributions to the development of modern social networks, and complements the evolutionary context of count models from "simple to complex". It uses the negative binomial distribution to break through duality and the zero-inflated model to break through non-correlation. The main contribution of this study is to fill the above gap, linking the strengths, weaknesses, and mathematical principles of each model to the specific complexity of social networks, clarifying when and why each model is suitable for our use in real life, and also constructing mathematical thinking logic and inference theory. This not only enriches the theoretical background of statistical methods in social network analysis but also provides practical guidance for model selection in real scenarios.

This paper starts by going over the basic math behind the Bernoulli, Poisson, negative binomial, and zero-inflated models. Then, it explains the missing logics and their own shortcomings of the Bernoulli distribution and Poisson distribution in the complex analysis of social networks, and analyzes their duality and the distribution shortcomings of traditional laws. The following parts will analyze how the negative binomial distribution breaks through duality, changing traditional social relationships to more complex directional relationships, and how the zero-inflated model breaks through non-correlation, regarding each person as a completely different individual. Finally, in the conclusion, it will close the loop on the hypothesis relaxation trajectory of the whole process, extract a universal framework, and look forward to cutting-edge directions such as the deep embedding of network structures.

This paper aims to address which types of complexity in social networks the mathematical mechanisms of different models correspond to solving, and how to connect the evolutionary logic of models through the "assumption relaxation trajectory". This study addresses the flaws of traditional models in social networks. It also fills prior research gaps by linking each model's mathematical logic to real social network needs. Additionally, it connects model evolution via the "assumption relaxation trajectory" and provides practical guidance for model selection, supporting both academic and real-world analysis.

2.Literature review

Binary variables in a sequence are usually considered independent when dealing with typical Bernoulli distributions, but this assumption often doesn't hold up when looking at real data. If trials are related in a time series or success chances vary, overdispersion will exist. A more detailed version of the Generalized Bernoulli Process (GBP) was mentioned to address this. It makes it possible for binary variables to influence each other by relaxing the idea that they have to be

DOI: 10.54254/2753-8818/2026.HZ27624

independent [1]. The traditional social network analysis depends on the Bernoulli model, which cannot capture linked interactions in social networks, like when users' actions are related. GBP is better at this, but it only works for two simple outcomes (yes and no) and struggles with count data, like how many times people interact.

Building on this, using the Poisson distribution instead of GBP works for count data is better, but the count data of social networks generally exhibits heavy-tailed characteristics because of a small number of users engaging in high-frequency interactions, and the majority of users only participate in low-frequency interactions. This heavy-tailed nature leads to a prominent overdispersion phenomenon: the variance of interaction counts is much larger than the mean, and the probability of extreme high-frequency users appearing is far higher than the prediction of ordinary distributions.

The Poisson distribution, which relies on the assumption that "mean equals variance" and has a rapidly decaying tail, is completely unable to adapt to heavy-tailed characteristics. Recent extensions of the Poisson family for asymmetric count data (e.g., weighted/exponential-mixture approaches) address overdispersion to some extent, but still cannot explain excess zeros [2]. Since zeros might come from users who are inactive structural zeros or just not active by chance stochastic zeros, the probability of zero values occurring in social networks is very high. If zero values are not taken into account, the results obtained will have a large error

To better handle overdispersion in count data, the negative binomial (NB) distribution is a better choice. The negative binomial (NB) distribution effectively addresses the overdispersion caused by heavy tails by introducing a dispersion parameter to flexibly adjust the relationship between variance and mean. It doesn't require the mean and variance to be equal, so it works well for explaining strong overdispersion in social networks [3]. For example, most users interact rarely, but a few interact a lot in social networks. But NB has a problem; it treats all zeros the same and can't tell the difference between structural and stochastic zeros. This is a gap in social network analysis. Zeros can come from different reasons, like a new user who hasn't interacted yet, versus an active user who just didn't interact once.

This pattern shows a clear trend: each model fixes some problems of the one before it, but faces new challenges with more complex questions. The traditional Bernoulli model fails with linked data; GBP fixes that, but not count data issues; Poisson works for counts, but not overdispersion; PEBWEB is better with overdispersion, but not with different kinds of zeros. NB handles overdispersion but not different types of zeros. The Zero-Inflated Negative Binomial (ZINB) model, on the other hand, covers those missing pieces. It clearly splits zeros into structural permanent absence and stochastic temporary absence while keeping NB's ability to handle overdispersion [3]. This makes ZINB better for complex social networks, where interactions have both overdispersion and different kinds of zeros.

So, the move from simpler models to more complex ones—Bernoulli to GBP to Poisson to PEBWEB to NB to ZINB—shows a clear effort to match how social networks really work. Every new model improves on the limits of the one before. More complex models like ZINB have fewer rules and are more accurate because they account for the many sides of real interactions. As social networks get more complex, this trend shows why models that are both flexible and realistic are more and more needed for useful analysis.

Most existing reviews analyze social network analysis models in isolation. They neither systematically distinguish the practical application scenarios and the advantages or disadvantages of each model, nor do they mathematically verify the differences between these models. Therefore, this paper aims to analyze and summarize the application of different models in social network analysis

DOI: 10.54254/2753-8818/2026.HZ27624

and to provide guidance for practitioners and researchers in the social network industry regarding model analysis and selection in the future.

3.Methodology

The keywords such as "social network", "Bernoulli random distribution", "Poisson random distribution", and "Zero-Inflated model" were used to search on platforms like Google Scholar. Articles published between 2015 and 2025 were all included in these statistics. Since this paper aims to discuss the uses, advantages, and disadvantages of several models and distributions in analyzing social networks, papers that focus entirely on practice and lack theoretical contributions were excluded. This paper also analyzed each cited literature in terms of accuracy, appropriateness, application, and scope.

4.Results

Table 1. Comparison between four distributions

Model	Key characteristics	Advantages	Limitations	Applicable Scenarios	Diagnostic Signals	Recommended Tests
Bernoulli Distribution	Binary outcomes are independent (no dependence between observations).	Easy to interpret for binary data.	Cannot explain dependent data and overdispersion.	Simple networks with independent binary tie.	Data is strictly binary (only 0/1, no counts); observations show no interdependence.	Chi-square goodness-of-fit test, Check for only two outcome categories.
Poisson Distribution	Mean of counts equals their variance.	Naturally models count data, consider some rare-happened situations and sparse network.	Variance equals to mean cannot easily happened in real life.	Networks with stable count data where mean ≈ variance.	Count data with mean approximately equal to variance; no obvious overdispersion.	Mean-variance ratio test.
Negative Binomial Distribution (NB)	Variance can exceed mean (accommodates overdispersion in count data).	Effectively handles overdispersion (variance > mean) in count data.	Cannot explain heterogeneous zero value.	Networks with severe overdispersion but few random zeros.	Count data with variance significantly greater than mean; zeros are mostly randomly generated (no obvious “structural zero” clustering).	Likelihood ratio test (compare fit of NB and Poisson models).
Zero-Inflated Negative Binomial (ZINB)	Zeros split into “structural zeros” (inherent absence) and “random zeros” (stochastic absence).	Addresses both overdispersion and Zero-inflation problems.	Higher computational complexity.	Complex networks with heterogeneous zeros and overdispersion.	Count data with variance > mean (overdispersion) and a large number of zero.	Vuong test (compare goodness-of-fit between ZINB and NB to determine if zero-inflation is needed).

As shown in Table 1, these four distributions each have some advantages and limitations, but from top to bottom, their complexity is increasing. From the Bernoulli distribution to ZINB, the more complex they are, the closer they get to real-life situations. When analyzing social networks,

DOI: 10.54254/2753-8818/2026.HZ27624

they become more accurate. So, the more complex the model is, the fewer restrictions there are on its application scenarios.

5.Discussion

With the exponential growth in the number of users and data volume on platforms such as Twitter, WeChat, and Facebook, social network data centered on interpersonal interaction patterns like message sending frequency, retweet counts, and relationship formation has become a focus of academic research. However, such data has unique statistical characteristics like excessive zeros (most user pairs never interact), overdispersion (variance is much larger than the mean), and complex generation mechanisms (zeros stem from both structural barriers and random inaction). These characteristics pose challenges to traditional statistical models. This section reviews the evolution of count models applied to social networks, focusing on their application scenarios, limitations, and latest progress.

Early studies used the simplicity of the Bernoulli distribution to model the existence of relationships, using "1" to indicate the presence of a connection and "0" to indicate absence. Dyadic connections are the basic logic that can analyze social networks, and as the social network assumption holds, individuals are not isolated [4]. This method actually nailed the simple yes-or-no kind of social ties, making it easier to start analyzing basic social networks. The Bernoulli distribution is used to describe a single random trial with only two possible outcomes (success and failure), and the sum of the probabilities of these two possible outcomes is 1. X can be a Bernoulli random variable. However, as research shifted to interaction intensity, its limitations became apparent: the Bernoulli framework simplifies relationships to a binary state and cannot quantify subtle dynamics such as the number of user interactions, the intensity of repeated communication, or zeros representing impossible connections.

To address the inability of the Bernoulli distribution to handle count data, researchers turned to the Poisson distribution, which models the number of events in a fixed interval using a single parameter λ under the assumption that "mean equals variance." Let X be a Poisson random variable; then the probability mass function (PMF) of X is

$$P_x(k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0,\ 1,\ 2,\ \ldots \tag{1}$$

However, through calculations, for a random variable $X \sim \text{Poisson}(\lambda)$ following a Poisson distribution, its mean is

$$E(X) = \sum_{k=0}^{\infty} k \cdot P(X = k) \tag{2}$$

Substituting into equation (1) and simplifying (by variable substitution m = k - 1 and using the Taylor expansion of the exponential function), $\mathrm{E}(\mathrm{X}) = \lambda$ can be obtained . Then, using the variance formula

$$Var(X) = E\left(X^2\right) - [E(X)]^2 \tag{3}$$

and deriving with the identity $\mathrm{X}^2 = \mathrm{X}(\mathrm{X}-1) + \mathrm{X}$, $\mathrm{E}\left(\mathrm{X}^2\right) = \lambda^2 + \lambda$ can be evaluated and further find that $\mathrm{Var}(\mathrm{X}) = \lambda$. In summary, the Poisson distribution satisfies [5]

DOI: 10.54254/2753-8818/2026.HZ27624

$$E(X) = Var(X) = \lambda \tag{4}$$

Yet, empirical studies have repeatedly revealed a key flaw: social network data violates Poisson's equidispersion assumption. When variance is greater than the mean,

$$\widetilde{E}_i[d] = E_i[d] + \frac{Vari[d]}{E_i[d]} \tag{5}$$

In a random network model, the expected degree of neighbors is equal to the overall average degree plus a factor related to the variance of the degree distribution and the average degree. When $\mathrm{Var_i}[\mathrm{d}] = 0$ (that is, all nodes are perfectly positively matched and the network is a regular network), the expected degree of neighbors is equal to the degree of the node itself. But in real-life relationships, there are differences in the degree distribution, and the variance will not be zero but greater than zero. Therefore, in this real-world example of the friendship paradox, there is a phenomenon that people's friends have, on average, more friends than themselves, and thus deduce that the expected degree of neighbors will be greater than the overall average degree, and the variance will not be equal to the mean. Proving that the variance of degree distributions in social networks is not equal to the mean [6].

The negative binomial (NB) distribution specifically addresses the shortcomings of the Poisson distribution. Zero-inflated and hurdle variants are reviewed in [7]. The basic logic of NB is to treat the mean parameter μ of the Poisson distribution as a random variable (following a gamma distribution), forming a "Poisson-gamma mixture distribution", and finally let the mean μ of the Poisson distribution follow a gamma distribution $\mu \sim \mathrm{Gamma}(\mathrm{r}, \beta)$. Its probability density function (PDF) is $\mathrm{f}(\mu) = \frac{\beta^{\mathrm{r}}}{\Gamma(\mathrm{r})}\mu^{\mathrm{r}-1}\mathrm{e}^{-\beta\mu}, \quad \mu > 0$. Among them, $\mathrm{r} > 0$ is the shape parameter, $\beta > 0$ the rate parameter, and the expectation and variance of the gamma distribution, which satisfy:

$$E(\mu) = \frac{r}{\beta}, \quad \mathrm{Var}(\mu) = \frac{r}{\beta^2} \tag{6}$$

Using the law of total expectation $\mathrm{E}(\mathrm{Y}) = \mathrm{E}[\mathrm{E}(\mathrm{Y}\mu)]$, the inner expectation can be seen as the expectation of the Poisson distribution $\mathrm{E}(\mathrm{Y}\mu_1) = \mu_1$ (mean = variance). The Outer expectations the expectation of a gamma distribution $\mathrm{E}(\mu_1) = \frac{\mathrm{r}}{\beta} = \mu$. Thus, the expectation of the negative binomial distribution is

$$E(Y) = \mu \tag{7}$$

Using the variance decomposition formula $\mathrm{Var}(\mathrm{Y}) = \mathrm{E}[\mathrm{Var}(\mathrm{Y}\mu)] + \mathrm{Var}[\mathrm{E}(\mathrm{Y}\mu)]$. The first term can be estimated as $\mathrm{E}[\mathrm{Var}(\mathrm{Y}\mu)] = \mathrm{E}(\mu) = \mu$. Combining $\mathrm{r} = \alpha^{-1}$ and $E(\mu) = \mu = \frac{r}{\beta}$, the second term can be estimated as $\mathrm{Var}(\mu) = \frac{r}{\beta^2} = \frac{\alpha^{-1}}{\left(\frac{1}{\alpha\mu}\right)^2} = \alpha\mu^2$. So the variance of the negative binomial distribution can be written as

$$\mathrm{Var}(Y) = E[\mathrm{Var}(Y\mu)] + \mathrm{Var}[E(Y\mu)] = \mu + \alpha\mu^2 \tag{8}$$

It explicitly relaxes the "mean equals variance" constraint by introducing a dispersion parameter α, with the variance expressed as "variance = mean + α×mean²." α quantifies the degree of

DOI: 10.54254/2753-8818/2026.HZ27624

overdispersion in count data: when $\alpha > 0$, variance exceeds the mean, allowing it to accurately capture the unbalanced pattern in data where "a few users interact frequently while most interact infrequently." The larger α is, the better it fits data with important overdispersion, such as interaction counts in social networks [8]. Thus, the negative binomial distribution is more consistent with empirical data than the Poisson distribution.

However, while it resolves overdispersion, the negative binomial distribution fails to explicitly distinguish the generation mechanisms of zeros: it treats all zeros (including "never interacted" and "occasionally not interacted") as random fluctuations, which misaligns with the complexity of zeros in social networks.

The basic logic of zero-inflated models lies in distinguishing two types of zeros: structural zeros (interactions that will never occur due to objective constraints, such as between strangers with no connection, belonging to "impossible" zeros) and sampling zeros (interactions that could occur but do not, such as occasional non-interaction between friends, belonging to "randomly non-occurring" zeros). Zero-Inflated Models (ZIM) are the mainstream method for such data, but although the traditional ZIP can handle excessive zeros, its basic logic is Poisson random variables, which means it assumes that the mean of the count component is equal to the variance, and thus cannot cope with overdispersion caused by population heterogeneity and omitted covariates. However, the negative binomial distribution allows the variance to vary independently of the mean by introducing a dispersion parameter r. Therefore, combining the negative binomial distribution as the count component with the zero-inflated structure forms the Zero-Inflated Negative Binomial model (ZINB), which can simultaneously solve the problems of "excessive zeros" and "overdispersion".

The Zero-Inflated Negative Binomial (ZINB) model separates the sources of zero values and solves overdispersion through a mixed structure consisting of a zero-inflated component and a negative binomial count component. Its probability mass function (PMF) is as follows

$$P(Y_i = y_i) = \pi_i + (1 - \pi_i) \cdot \left(\frac{r}{\mu_i + r}\right)^r \quad \text{if } y_i = 0 \tag{9}$$

$$P(Y_i = y_i) = (1 - \pi_i) \cdot \frac{\Gamma(y_i + r)}{\Gamma(r)\, y_i!} \cdot \left(\frac{\mu_i}{\mu_i + r}\right)^{y_i} \cdot \left(\frac{r}{\mu_i + r}\right)^r \quad \text{if } y_i > 0 \tag{10}$$

These two formulas are from [7]. The zero-inflated component π_{i} represents the probability of structural zeros. When $\mathrm{y}_{\mathrm{i}} = 0$, the probability of zero values consists of π_{i} (structural zeros) and $(1 - \pi_{\mathrm{i}}) \cdot \left(\frac{\mathrm{r}}{\mu_{\mathrm{i}} + \mathrm{r}}\right)^{\mathrm{r}}$ (sampling zeros). Negative binomial count component is for non-zero counts non-structural zeros, μ_{i} is the mean of the negative binomial distribution (the average rate of event occurrence); r is the dispersion parameter (controlling the degree of overdispersion; the smaller r is, the more serious the overdispersion is); $\Gamma(\cdot)$ is the gamma function, which connects discrete counts with continuous distributions to ensure the rationality of probability calculations.

The mean and variance formulas of the ZINB model demonstrate its advantage in handling overdispersion.

$$E(Y_i) = (1 - \pi_i) \cdot \mu_i \tag{11}$$

Only individuals with non-structural zeros may experience events, so the mean is jointly determined by the proportion of the non-zero population and their average count μ_{i}.

DOI: 10.54254/2753-8818/2026.HZ27624

$$\mathrm{Var}\left(Y_i\right)=\left(1-\pi_i\right)\cdot\mu_i\cdot\left(1+\frac{\mu_i}{r}+\pi_i\cdot\mu_i\right) \tag{12}$$

$\frac{\mu_i}{r}$ derived from the overdispersion characteristic of the negative binomial distribution (the smaller r is, the larger this term is, and the more the variance exceeds the mean). The $\pi_i\cdot\mu_i$ was derived from the zero-inflated structure (the higher the probability of structural zeros, the larger this term is, the larger the variance is). Together, they enable the variance to be significantly larger than the mean, perfectly adapting to overdispersal data [7].

Therefore, in the future, when analyzing simple binary problems, the Bernoulli distribution can be used to analyze social networks merely as a simple independent yes-or-no relationship. However, when analyzing real social networks, the relationships between many users are multi-frequency. Thus, the Poisson distribution can be better used to model the number of events occurring within a certain time or space, so as to describe the differences in the activity of relationships between different nodes in social networks. Yet, in more complex and real social networks, the Poisson distribution, which assumes the ideal discrete situation where the mean is equal to the variance, cannot analyze the data well and will lead to large errors. Therefore, the negative binomial distribution introduces an additional parameter to allow the variance to be greater than the mean, which is more in line with the situation in real life, where a small number of groups interact frequently while the majority interact infrequently. On the basis of the negative binomial distribution, the zero-inflated negative binomial distribution can be used to make up for the defect of the negative binomial distribution that assumes all zero values occur randomly. It takes into account the large number of zero values and a small number of non-zero data in social network analysis and analyzes the influencing factor of structural zero values. So, when analyzing social networks in the future, it is very necessary to consider more complex models. This paper, starting from the explanation of simple models to more complex ones, intends to analyze the complexity of real social networks and the logical direction for future analysis of social networks. It is hoped that the research on social network analysis can develop more in-depth, moving towards more complex models and logics. The ZINB also has drawbacks in the study of complex social network analysis, which need to be considered in future research. If the research sample size is small, the ZINB model will be unable to stably estimate these parameters due to insufficient data, leading to biased or excessively volatile parameter estimates and failing to reflect the real social network patterns. Moreover, when the proportion of zero values in the data is extremely low (such as <5%) or extremely high (such as >95%), the "zero-inflation parameter" and the "dispersion parameter of the negative binomial component" will become difficult to distinguish. For example, if there is almost no user interaction on a certain social platform, the model cannot determine whether the zero values result from structural non-interaction or simply random sparsity.

Therefore, when selecting a model for social network analysis, the first step is to determine whether the data is binary—i.e., whether the relationships between social network users are simply "yes or no." If the data is ordinary binary data, the Bernoulli distribution is used for analysis. However, if the data is not binary but complex count data, like interaction frequency and connection count, it is necessary to consider whether the data is overdispersed and the number of zero values.

$$\mathrm{VMR}=\frac{\sigma^2}{\mu} \tag{13}$$

DOI: 10.54254/2753-8818/2026.HZ27624

Specifically, for count data, if there is no overdispersion (Variance-to-Mean Ratio, VMR ≤ 1) and the number of zero values is small (no zero-inflation), the Poisson distribution is used; if there is no overdispersion (VMR ≤ 1) but the number of zero values is relatively large (with zero-inflation), the Zero-Inflated Poisson (ZIP) model is used; if there is overdispersion (VMR > 1) but no zero-inflation, the NB distribution is selected; if there is both overdispersion (VMR > 1) and zero-inflation, the ZINB model is selected.

6.Conclusion

This study focuses on the main data characteristics of social networks, such as the flood of zero values, overdispersion, and complex social relationships. It reveals the limitations of traditional models and concludes that, from the basic Bernoulli distribution to the Poisson distribution, then to the negative binomial distribution, and further to the zero-inflated negative binomial distribution, as the complexity of the model increases, it becomes more suitable for the analysis and calculation of social networks in real life. This paper explains from simple models to complex models step by step: starting with explaining that the Bernoulli distribution can only depict the duality of "presence or absence of interaction", then pointing out the contradiction between the Poisson distribution's assumption of "equality of mean and variance" and the strong overdispersion of real data, and then by introducing the negative binomial distribution, which relaxes the constraint of "mean equals variance" with a dispersion parameter to adapt to the overdispersion situation. The zero-inflated model helps distinguish between structural zeros and random zeros. This is useful because the negative binomial distribution doesn’t do a good job explaining zeros. It also emphasizes how differences in zero counts matter in real social networks. Plus, it allows a comparison between simple and more complex models when studying these networks.

This study takes a different approach by looking at models together instead of studying only one of them when examining social networks. It systematically establishes the mapping between the characteristics, logics of the Bernoulli, Poisson, negative binomial, and zero-inflated models and social scenarios, and clarifies the adaptation rules: "Bernoulli for simple traditional analysis, Poisson for non-binary analysis, negative binomial for overdispersion, and zero-inflated for zero-value heterogeneity". At the same time, this paper also solves the problem of blind selection of models in social network research. It uses mathematical derivation and explanation from binary assumptions to multi-constraint breakthroughs, providing a basis for selecting applicable models for complex social network data.

For future research, this study provides a framework for model selection. If future studies aim to explain simple binary relationships in a small amount of data, the Bernoulli distribution can be used; if explaining unordered relationships within a certain period, the Poisson distribution is applicable; if focusing on dynamic social networks that are close to reality, the negative binomial distribution or the zero-inflated negative binomial model can be selected according to the explanations provided above. Academic researchers can choose models suitable for their research based on the evolution laws of these four models; practitioners in social platforms can accurately identify "inert users" and "temporarily silent groups" through the zero-inflated model and formulate differentiated strategies.

This study has limitations in covering social network scenarios. It only focuses on static social network analysis, does not consider temporal dynamics, nor expands to multi-modal networks, and fails to take into account the impact of emotions in real life on relationships and other complex factors that can affect relationships. Meanwhile, when discussing the zero-inflated model, this paper does not explain the situation where the number of zero values in the data is less than expected, and

DOI: 10.54254/2753-8818/2026.HZ27624

the situation where the difference in the generation mechanisms of structural zeros and sampling zeros is small.

Future research can introduce factors such as periodic factors to develop dynamic negative binomial models to capture the temporal fluctuations of interaction frequencies, and time-varying zero-inflated models to distinguish the temporal characteristics of long-term inertia and short-term silence. It can also study the differences in the generation mechanisms of structural zeros and sampling zeros to adapt to the dynamic evolution of social networks. In addition, static networks can be expanded to large-scale dynamic networks, and graph theory methods can be combined to optimize computational efficiency, support real-time interaction analysis, and solve the problem of complex social network analysis.

References

[1] Lee J. Generalized Bernoulli process with long-range dependence and fractional binomial distribution. Dependence Modeling. 2021; 9(1): 1–12. doi: 10.1515/demo-2021-0100.

[2] Alomair A, Ahsan-ul-Haq M. A new extension of Poisson distribution for asymmetric count data: theory, classical and Bayesian estimation with application to lifetime data. PeerJ Computer Science. 2023; 9: e1748. doi: 10.7717/peerj-cs.1748.

[3] Stoklosa J, Blakey RV, Hui FKC. An overview of modern applications of negative binomial modelling in ecology and biodiversity. Diversity. 2022; 14(5): 320. doi: 10.3390/d14050320.

[4] Brass DJ. New developments in social network analysis. Annual Review of Organizational Psychology and Organizational Behavior. 2022; 9: 225–246. doi: 10.1146/annurev-orgpsych-012420-090628.

[5] Chan SH. Introduction to Probability for Data Science. Ann Arbor: Michigan Publishing; 2021. ISBN 978-1-60785-747-1 (ebook).

[6] Jackson MO. The friendship paradox and systematic biases in perceptions and social norms. arXiv preprint arXiv: 1605.04470; 2017 (v6).

[7] Feng CX. A comparison of zero-inflated and hurdle models for modeling zero-inflated count data. Journal of Statistical Distributions and Applications. 2021; 8: 8. doi: 10.1186/s40488-021-00121-4.

[8] NCSS, LLC. Negative Binomial Regression (Chapter 326). In: NCSS Statistical Software Documentation. Kaysville, UT: NCSS, LLC; 2025. (Chapter PDF).