

BISoN: A Bayesian framework for inference of social networks

Jordan Hart*

Department of Psychology
University of Exeter
Exeter, UK

jordan.da.hart@gmail.com
[ORCID](#)

Nguyễn Nhật Hoàng(**)

Faculty of Information Science and
Engineering
University of Information Technology
HCM city, VN
20520516@gm.uit.edu.vn

Abstract— Bài báo giới thiệu BISoN, một khuôn khổ Bayesian thống nhất để mô hình hóa mạng lưới xã hội từ dữ liệu quan sát. BISoN giúp ước lượng sự không chắc chắn trong trọng số các cạnh mạng, xử lý các yếu tố gây nhiễu và mô hình hóa ảnh hưởng ở mức độ quan sát, đồng thời tương thích với các phương pháp phân tích mạng lưới xã hội phổ biến. Khuôn khổ này có thể áp dụng cho các loại dữ liệu quan sát xã hội như dữ liệu nhị phân, số lượng và thời gian, đồng thời cho phép thực hiện các phân tích thống kê như kiểm tra mối quan hệ không ngẫu nhiên và hồi quy các đặc tính mạng. BISoN giúp tăng cường độ tin cậy của các suy luận khoa học và mở rộng khả năng kiểm tra các giả thuyết mới. Các công cụ và mã nguồn R đã được cung cấp để hỗ trợ áp dụng khuôn khổ này.

Keywords— animal social network analysis, Bayesian inference, dyadic regression, network metrics, nodal regression

(*): Đại diện cho nhóm tác giả bài báo

(**): Người viết báo cáo

I. GIỚI THIỆU

Phân tích mạng xã hội được cho là một trong những khuôn khổ phổ biến nhất trong nghiên cứu về tính xã hội (Croft và cộng sự, 2016). Trong nhiều nghiên cứu khoa học, dữ liệu được thu thập ảnh hưởng đến cả việc giải thích các phân tích mạng xã hội và độ chính xác của chúng (Whitehead, 2008a).

Các mạng lưới thường được xây dựng bằng cách sử dụng một biện pháp chuẩn hóa về tính xã hội, chẳng hạn như tỷ lệ thời gian lấy mẫu mà mỗi cặp dành để tham gia vào một hành vi xã hội (ví dụ: Chỉ số tỷ lệ đơn giản [SRI]; Cairns & Schwager, 1987). Các biện pháp chuẩn hóa này đảm bảo rằng các cặp được quan sát trong thời gian dài hơn không bị xác định sai là có tính xã hội hơn (Whitehead, 2008a). Điều này rất quan trọng vì việc quan sát các sự kiện xã hội có thể là một thách thức và việc lấy mẫu đồng đều trên tất cả các cặp không phải lúc nào cũng khả thi (Croft và cộng sự, 2008). Mặc dù các biện pháp chuẩn hóa về cấu trúc xã hội sẽ không bị ảnh hưởng bởi thời gian lấy mẫu, nhưng chúng sẽ đi kèm với các mức độ chắc chắn khác nhau. Ví dụ, trọng số cạnh sẽ được coi là 0,5 nhất định cho cả trường hợp các cá nhân được nhìn thấy cùng nhau một lần và xa nhau một lần, và tương tự như vậy khi các cá nhân đã ở cùng nhau 100 lần và xa nhau 100 lần. Các giải pháp hiện có cho vấn đề này nhằm mục đích định lượng trước độ chính xác và sức mạnh của mạng xã hội (Hart và cộng sự, 2021; Whitehead, 2008b), lọc ra các điểm dữ liệu có độ không chắc chắn cao (James và cộng sự, 2009) hoặc đưa nỗ lực lấy mẫu vào các thuật ngữ trong các mô hình thống kê (Franks và cộng sự, 2021). Các giải pháp hiện có này có thể hữu ích, nhưng cũng không tránh khỏi việc mất thông tin và có thể đưa ra sai lệch trong các phân tích. Mặc dù sự không chắc chắn có thể xuất hiện trong hầu hết các phép đo (ví dụ: tuổi, kích thước hoặc thậm chí là giới tính), các nghiên cứu

được đề cập ở trên đã nhấn mạnh rằng sự không chắc chắn do nỗ lực lấy mẫu có thể lớn hơn nhiều lần so với các nguồn không chắc chắn khác về phép đo. Bất kỳ nguồn không chắc chắn chính nào cũng có thể ảnh hưởng đáng kể đến hiệu suất của các mô hình thống kê, do đó, điều quan trọng là phải tính đến sự không chắc chắn do nỗ lực lấy mẫu một cách thích hợp (Wasserman, 2004). Mặc dù vậy, các phương pháp để ước tính đầy đủ và truyền bá sự không chắc chắn thông qua các phân tích mạng xã hội vẫn chưa được phát triển đầy đủ.

Trong bài báo này, nhóm tác giả giới thiệu BISoN: một khuôn khổ Bayesian thống nhất chung để mô hình hóa dữ liệu mạng xã hội. BISoN nắm bắt sự không chắc chắn trong trọng số cạnh, tận dụng tối đa dữ liệu có sẵn; truyền bá sự không chắc chắn thông qua các phân tích hạ lưu; và có thể kiểm soát các hiệu ứng xã hội và phi xã hội ở mọi cấp độ (cá nhân, cặp đôi, nhóm, quan sát, v.v.). Bất kỳ loại phân tích mạng xã hội nào cũng có thể được thực hiện trong khuôn khổ của nhóm tác giả, bao gồm hồi quy cặp đôi và nút, các bài kiểm tra trọng số cạnh không ngẫu nhiên và ước tính các thuộc tính mạng có cấu trúc. Khuôn khổ BISoN bao gồm (a) một mô hình trọng số cạnh, phụ thuộc vào loại dữ liệu, xây dựng các trọng số cạnh và mạng có sự không chắc chắn từ dữ liệu thực nghiệm và (b) các phân tích hạ lưu sử dụng các trọng số cạnh ước tính để tính đến sự không chắc chắn trong mạng. Bản thảo này được coi là bản tổng quan kỹ thuật về BISoN và giả định rằng bạn đã quen thuộc với các khái niệm chính của thống kê Bayesian và phân tích mạng xã hội. Những độc giả không quen thuộc với quy trình thống kê Bayesian và phân tích mạng xã hội được hướng dẫn đến các văn bản giới thiệu như Kruschke (2015), McElreath (2020), Carrington và cộng sự (2005) và Croft và cộng sự (2008). Nhóm tác giả sẽ tập trung vào cách khuôn khổ của nhóm tác giả có thể được áp dụng cho các hệ thống động vật nhưng nguyên tắc cơ bản có thể được áp dụng cho bất kỳ loại phân tích mạng nào trong đó các cạnh mạng được suy ra. Khuôn khổ BISoN mang đến cơ hội tạo ra suy luận khoa học đáng tin cậy, linh hoạt và phong phú trong nghiên cứu các hệ thống xã hội.

II. BISoN — PHÂN TÍCH BAYESIAN CHO MẠNG LƯỚI XÃ HỘI

Phần này trình bày cách BISoN (Bayesian Inference of Social Networks) sử dụng phương pháp Bayesian để mô hình hóa mạng lưới xã hội từ dữ liệu quan sát. Đặc biệt, BISoN tập trung vào **mô hình trọng số cạnh** (edge weight models) cho các loại dữ liệu khác nhau, bao gồm **dữ liệu nhị phân**, **dữ liệu đếm** và **dữ liệu thời gian**.

A. Đo lường sự không chắc chắn trong trọng số các cạnh (Edge Weights)

BISoN sử dụng phân tích Bayesian để mô hình hóa sự không chắc chắn trong trọng số các cạnh của mạng lưới xã hội. Điều

| Data type | Description | Edge interpretation |
|-----------|---|---|
| Binary | The presence or absence of a social event for each dyad is recorded in each sampling period | Probability of (or proportion of time) engaging in a social event in a fixed period of time |
| Count | The number of social events for each dyad is recorded in each sampling period | Rate of occurrence of social events per unit time |
| Duration | The amount of time each dyad spends engaged in a social event is recorded in each sampling period | Proportion of time spent engaging in a social event |

TABLE 1 Definitions of the three types of social observational data we have developed models for. Examples of standard edge weight interpretations are included in the third column.

này có nghĩa là **trọng số của mỗi cạnh** giữa hai cá thể trong mạng lưới không chỉ là một giá trị duy nhất mà là một **phân phối xác suất**, phản ánh mức độ không chắc chắn trong ước lượng.

1) Mô hình trọng số cạnh: Dữ liệu nhị phân

Dữ liệu nhị phân trong phân tích mạng lưới xã hội ghi lại sự có mặt hoặc vắng mặt của một sự kiện xã hội giữa hai cá thể trong mỗi khoảng thời gian quan sát. Mô hình trọng số cạnh ω_{ij} cho dữ liệu nhị phân trong **BISoN** được mô tả bằng **phân phối Bernoulli**, với trọng số cạnh chỉ có hai giá trị: 1 (có tương tác xã hội) hoặc 0 (không có tương tác xã hội). Phân phối này được xác định bởi công thức:

$$P(\omega_{ij} = 1|\theta) = \frac{1}{1 + \exp(-\theta)},$$

Trong đó, θ là tham số mô hình xác định xác suất của sự kiện xã hội xảy ra giữa cá thể i và cá thể j . Tham số θ có thể được ước lượng từ dữ liệu quan sát, ví dụ như tỉ lệ xuất hiện của các sự kiện xã hội giữa hai cá thể trong một khoảng thời gian xác định.

BISoN sử dụng phương pháp **Bayesian** để ước lượng tham số θ , nghĩa là không chỉ đưa ra một giá trị đơn lẻ cho xác suất mà còn tính đến sự không chắc chắn của giá trị này thông qua phân phối xác suất. Phương pháp này cho phép mô hình hóa sự biến động và bất định trong các quan sát, đồng thời giúp tăng cường tính chính xác và độ tin cậy của các phân tích mạng lưới xã hội, vì nó cung cấp một phạm vi các giá trị khả thi thay vì chỉ một ước lượng duy nhất.

2) Mô hình trọng số cạnh: Dữ liệu đếm

Dữ liệu đếm được sử dụng để ghi lại số lần tương tác xã hội giữa hai cá thể trong một khoảng thời gian quan sát cố định. Trong khuôn khổ **BISoN**, trọng số cạnh ω_{ij} cho dữ liệu đếm được mô hình hóa bằng **phân phối Poisson**, một phân phối phổ biến để mô tả tần suất xảy ra của các sự kiện độc lập trong một khoảng thời gian. Phân phối này được định nghĩa bởi công thức:

$$P(\omega_{ij}|\lambda) = \frac{\lambda^{\omega_{ij}} e^{-\lambda}}{\omega_{ij}!},$$

trong đó λ là tham số thể hiện tần suất trung bình của sự kiện xã hội xảy ra giữa cá thể i và cá thể j . Tham số λ phản ánh mức độ tương tác xã hội và được ước lượng từ dữ liệu quan

sát. Mô hình này đặc biệt phù hợp với dữ liệu có tính chất đếm như số lần chạm, gọi, hoặc tương tác thể chất giữa các cá thể.

Để xử lý sự không chắc chắn trong dữ liệu, BISoN không chỉ ước lượng một giá trị cố định cho λ mà còn mô hình hóa nó như một **biến ngẫu nhiên**, sử dụng phương pháp Bayesian để tạo ra **phân phối posterior**. Phân phối posterior cho phép xác định mức độ tin cậy của các giá trị λ và phản ánh sự biến động trong các quan sát. Điều này giúp nâng cao độ chính xác và độ tin cậy của phân tích mạng lưới xã hội, đồng thời đảm bảo các kết luận khoa học không bị sai lệch do sự không chắc chắn trong dữ liệu.

3) Mô hình trọng số cạnh: Dữ liệu thời gian

Dữ liệu thời gian trong phân tích mạng lưới xã hội ghi lại khoảng thời gian mà hai cá thể tương tác với nhau trong một sự kiện xã hội, thay vì chỉ đếm số lần hoặc ghi nhận sự hiện diện hay vắng mặt của sự kiện. Mô hình trọng số cạnh ω_{ij} cho dữ liệu thời gian trong **BISoN** được mô tả bằng **phân phối mũ** (Exponential Distribution), một phân phối thường được sử dụng để mô hình hóa thời gian giữa các sự kiện trong các hệ thống ngẫu nhiên. Công thức phân phối mũ cho trọng số cạnh là:

$$P(\omega_{ij}|\theta) = \theta e^{-\theta\omega_{ij}},$$

Trong đó, θ là tham số mô hình, xác định **tốc độ** của sự kiện xảy ra giữa hai cá thể i và j . Tham số θ thể hiện khả năng cá thể này tham gia vào các sự kiện xã hội trong một khoảng thời gian cụ thể. Các tham số mô hình này được ước lượng từ dữ liệu quan sát, chẳng hạn như thời gian các cá thể tương tác hoặc tham gia vào các sự kiện.

Khi sử dụng phân tích **Bayesian**, BISoN không chỉ ước lượng tham số θ duy nhất mà còn mô hình hóa sự không chắc chắn trong các tham số này, tạo ra một phân phối xác suất cho θ . Điều này giúp mô hình hóa sự biến động trong thời gian tương tác xã hội và nâng cao độ chính xác của các phân tích mạng lưới xã hội, đặc biệt khi dữ liệu quan sát có sự không đồng đều hoặc có các yếu tố ngẫu nhiên khó kiểm soát.

4) Xử lý các yếu tố xã hội và phi xã hội

Trong **BISoN**, không chỉ dữ liệu xã hội mà còn cả các **yếu tố phi xã hội** (non-social effects) cũng được xem xét khi mô hình hóa trọng số các cạnh trong mạng lưới xã hội. Các yếu tố xã hội có thể bao gồm các đặc điểm như sự tương tác trực

tiếp giữa các cá thể, trong khi các yếu tố phi xã hội có thể là những yếu tố bên ngoài ảnh hưởng đến hành vi xã hội, chẳng hạn như điều kiện môi trường, các yếu tố sinh lý, hoặc các sự kiện ngoại vi không liên quan trực tiếp đến mạng xã hội nhưng lại có ảnh hưởng gián tiếp đến sự tương tác giữa các cá thể. BISO_N cho phép tích hợp các yếu tố này vào mô hình trọng số cạnh thông qua các tham số bổ sung, giúp cải thiện độ chính xác và sự linh hoạt của mô hình.

Khi mô hình hóa các yếu tố phi xã hội, BISO_N sử dụng các tham số điều chỉnh để phân biệt giữa ảnh hưởng của các yếu tố xã hội và phi xã hội. Việc này có thể được thực hiện bằng cách đưa vào các biến giải thích bổ sung trong mô hình Bayesian, từ đó cho phép **ước lượng trọng số cạnh** một cách chính xác hơn. Bằng cách này, BISO_N không chỉ giúp hiểu rõ hơn về mối quan hệ xã hội giữa các cá thể mà còn đưa ra những đánh giá chi tiết về những yếu tố bên ngoài có thể ảnh hưởng đến mạng lưới xã hội, đồng thời giúp giảm thiểu các yếu tố nhiễu trong quá trình phân tích.

III. PHÂN TÍCH TIẾP THEO VỚI SỰ KHÔNG CHẮC CHẮN

Phần này trình bày các phương pháp phân tích mạng lưới xã hội dựa trên các trọng số cạnh đã được ước lượng với sự không chắc chắn. Các phương pháp này đảm bảo rằng sự không chắc chắn được lan truyền qua toàn bộ quá trình phân tích.

A. Kiểm tra trọng số cạnh không ngẫu nhiên

Mục tiêu của phân tích kiểm tra trọng số cạnh không ngẫu nhiên là xác định liệu các trọng số cạnh trong mạng lưới xã hội có khác biệt một cách đáng kể so với mô hình ngẫu nhiên. Trong phân tích này, một mô hình ngẫu nhiên được sử dụng làm cơ sở tham chiếu, giả định rằng các trọng số cạnh giữa các cá thể là độc lập và phân phối đều. Sau đó, các trọng số cạnh quan sát được w_{ij} được so sánh với các giá trị kỳ vọng trong mô hình ngẫu nhiên. Nếu xác suất $P(w_{ij} > \text{ngưỡng})$ tính từ phân phối posterior của mô hình Bayesian cho thấy rằng trọng số cạnh vượt quá mức ngưỡng nhất định với xác suất cao, thì có thể kết luận rằng trọng số đó không phải là ngẫu nhiên. Điều này giúp xác định các mối quan hệ xã hội có cấu trúc, nơi các

cá thể có xu hướng tương tác nhiều hơn so với những gì được mong đợi trong một mạng lưới ngẫu nhiên. Phương pháp này giúp hiểu rõ hơn về các yếu tố xã hội và phi xã hội ảnh hưởng đến mạng lưới.

B. Hồi quy dyadic

Hồi quy dyadic trong BISO_N được sử dụng để phân tích các yếu tố ảnh hưởng đến trọng số cạnh giữa hai cá thể trong mạng lưới xã hội. Mô hình này cho phép xác định các biến giải thích (như đặc điểm cá nhân, sự gần gũi về địa lý hoặc hành vi xã hội) có ảnh hưởng đến mức độ tương tác giữa hai cá thể. Trọng số cạnh w_{ij} giữa cá thể i và j được mô hình hóa thông qua phân phối Poisson, với tham số λ_{ij} biểu thị tần suất của sự kiện xã hội xảy ra giữa hai cá thể. Mô hình hồi quy có thể được viết dưới dạng:

$$\log(\lambda_{ij}) = \beta_0 + \beta_1 x_{ij} + \beta_2 y_{ij} + \dots,$$

Trong đó x_{ij}, y_{ij} là các biến giải thích cho mỗi quan hệ giữa i và j , và các tham số β được ước lượng để chỉ ra mức độ ảnh hưởng của từng yếu tố. Hồi quy dyadic giúp hiểu rõ hơn về các yếu tố như sự tương tác xã hội, khoảng cách địa lý, hoặc các yếu tố cá nhân ảnh hưởng đến mức độ kết nối giữa hai cá thể trong mạng lưới. Phương pháp này không chỉ tính toán trọng số cạnh mà còn cho phép giải thích các yếu tố thúc đẩy hoặc hạn chế sự tương tác giữa các cá thể trong mạng lưới xã hội.

C. Hồi quy nodal

Hồi quy nodal trong BISO_N được sử dụng để phân tích các đặc điểm của từng cá thể trong mạng lưới xã hội dựa trên **độ trung tâm** (node centrality) của họ, tức là mức độ kết nối của một cá thể với các cá thể khác trong mạng. Độ trung tâm của cá thể i được tính bằng tổng trọng số của các cạnh kết nối với cá thể này, tức là tổng các trọng số cạnh w_{ij} với tất cả các cá thể j trong mạng lưới:

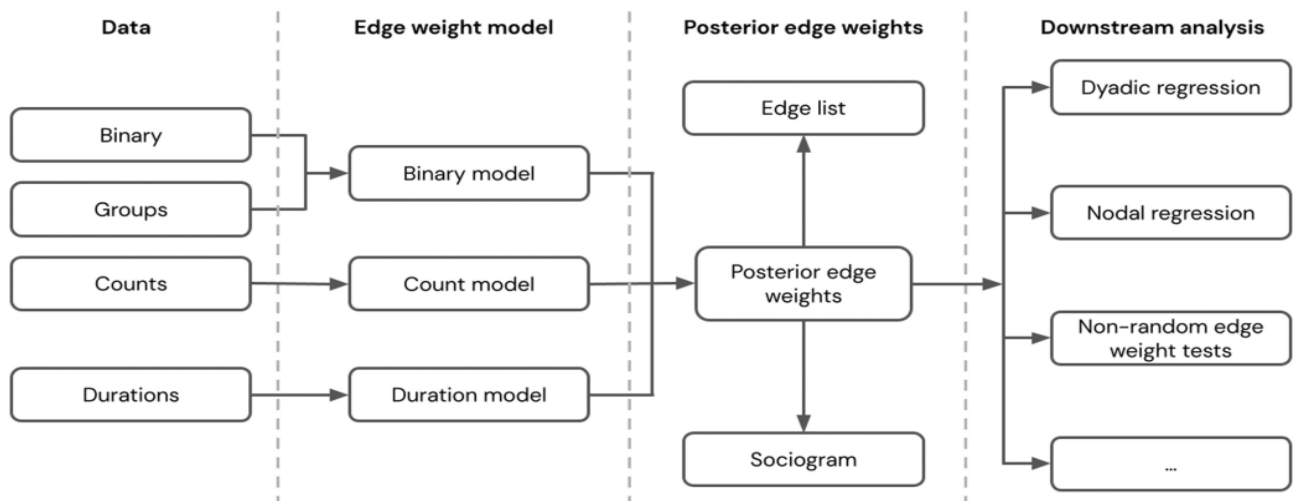


FIGURE 1 Tổng quan sơ đồ của khuôn khổ BISO_N. Các loại dữ liệu quan sát khác nhau ở các mức tổng hợp khác nhau được đưa vào mô hình trọng số cạnh thích hợp cho loại dữ liệu. Mô hình trọng số cạnh tạo ra các phân phối sau trên các trọng số cạnh, định lượng sự không chắc chắn trên chúng. Các trọng số cạnh sau có thể được trực quan hóa trong một biểu đồ xã hội với sự không chắc chắn hoặc được trích xuất dưới dạng danh sách cạnh với các khoảng tin cậy. Các trọng số cạnh sau cũng có thể được trích xuất cho các phân tích hạ lưu như phân tích hồi quy, kiểm tra trọng số cạnh không ngẫu nhiên và nhiều hơn nữa.

$$\text{Centrality}_i = \sum_j \omega_{ij}.$$

Hồi quy nodal sau đó phân tích mối quan hệ giữa độ trung tâm của cá thể và các yếu tố ảnh hưởng đến nó, như đặc điểm cá nhân (tuổi, giới tính, v.v.) hoặc các yếu tố xã hội khác. Mô hình hồi quy có thể được viết dưới dạng:

$$\text{Centrality}_i \sim \beta_0 + \beta_1 z_i + \dots,$$

Trong đó z_i là các biến giải thích (ví dụ, đặc điểm cá nhân của cá thể i), và các tham số β chỉ ra mức độ ảnh hưởng của từng yếu tố đến độ trung tâm của cá thể trong mạng lưới. Hồi quy nodal giúp xác định các cá thể có mức độ kết nối cao trong mạng lưới và hiểu được các yếu tố ảnh hưởng đến mức độ kết nối đó. Phương pháp này rất hữu ích để phân tích vai trò và vị trí của các cá thể trong mạng xã hội, đồng thời có thể giúp phát hiện các cá thể trung tâm hoặc quan trọng trong mạng lưới.

IV. VÍ DỤ VỚI DỮ LIỆU TỔNG HỢP

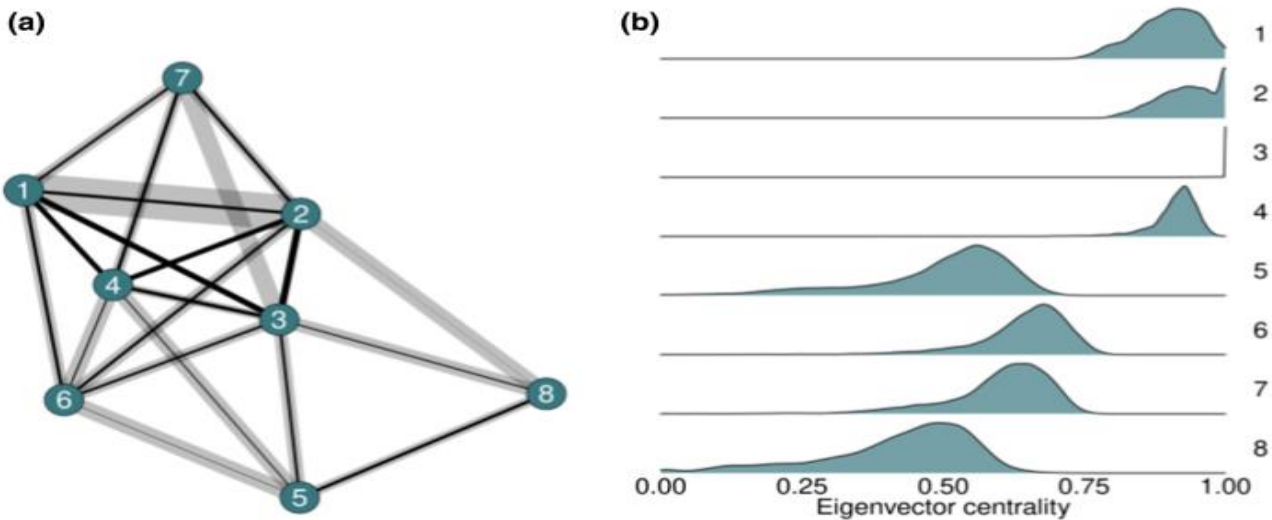
Bài báo trình bày một ví dụ dựa trên tập dữ liệu tổng hợp về các sự kiện xã hội mô phỏng dữ liệu đếm. Tập dữ liệu tổng hợp được tạo ra bằng cách gán trọng số cạnh không định hướng cơ bản cho 28 cặp đôi tương ứng với tất cả các cặp 8 cá nhân có thể sống trong một nhóm xã hội khép kín. Các cá nhân 1–4 được gán cho một điều kiện xử lý và các cá nhân 4–8 được gán cho một nhóm đối chứng. Các cặp đôi giữa hai cá nhân trong điều kiện xử lý được gán trọng số cạnh cao hơn các cặp đôi khác, để tạo ra hiệu ứng của điều kiện xử lý đối với tính trung tâm. Các sự kiện xã hội được mô phỏng từ một quá trình điểm Poisson dựa trên cả trọng số cạnh cơ bản và hiệu ứng phi xã hội ở cấp độ quan sát của vị trí. Có sáu vị trí có thể quan sát các sự kiện xã hội được sử dụng để mô phỏng hiệu ứng của các khả năng hiển thị khác nhau ở các vị trí khác nhau. Từ 10 đến 50 khoảng thời gian quan sát có độ dài cố định (đơn vị thời gian tùy ý) đã được mô phỏng cho mỗi cặp đôi. Bạn có thể tìm thấy thông tin chi tiết đầy đủ về mô phỏng trong mã ví dụ.

Bộ dữ liệu tổng hợp mô tả số lượng sự kiện cho mỗi giai đoạn quan sát, do đó mô hình thích hợp cho dữ liệu này là mô hình đếm. Mô hình đầy đủ được định nghĩa là

$$\begin{aligned} X_{ij}^{(n)} &\sim \text{Poisson}(\lambda_{ij}^{(n)} D_{ij}^{(n)}) \\ \log(\lambda_{ij}^{(n)}) &= \omega_{ij} + L_{ij}^{(n)} \\ \omega_{ij} &\sim \text{Normal}(0, 1) \\ L_{ij}^{(n)} &\sim \text{Normal}(0, \sigma_L^2) \\ \sigma_L &\sim \text{Half-Normal}(0, 1) \end{aligned}$$

trong đó $X_{ij}^{(n)}$ là số lượng sự kiện xã hội giữa i và j trong chu kỳ quan sát thứ n của chúng, $\lambda_{ij}^{(n)}$ là tỷ lệ sự kiện xã hội từ chu kỳ tương ứng, $D_{ij}^{(n)}$ là độ dài của chu kỳ lấy mẫu tương ứng (được cố định là 1 ở đây để đơn giản), $\omega_{ij}^{(n)}$ là trọng số cạnh cấp độ dyad và $L_{ij}^{(n)}$ là hiệu ứng cấp độ quan sát tương ứng với một trong sáu vị trí. Các phân phối trước được chọn bằng cách sử dụng kiểm tra dự đoán trước để đảm bảo các tham số chỉ mang các giá trị hợp lý về mặt sinh học. Mô hình được triển khai trong Stan và được lắp trong R, sử dụng gói Rstan (Nhóm cốt lõi R, 2020; Nhóm phát triển Stan, 2020). Các phương pháp tiếp cận khác như phép xấp xỉ Laplace lồng nhau tích hợp cũng có thể được sử dụng, hãy xem Thông tin hỗ trợ để thảo luận ngắn gọn về các thuật toán lắp ghép có sẵn (S6) (Rue và cộng sự, 2009).

Mức độ phù hợp của mô hình được kiểm tra trực quan bằng cách xem xét các chuỗi và tính toán chỉ số hội tụ Gelman-Rubin \hat{R} (Gelman & Rubin, 1992). Sau khi mức độ phù hợp của mô hình được xác minh, các ước lượng tham số của trọng số các cạnh ω_{ij} giữa các cặp cá thể (dyads) đã được trích xuất. Hình 2a hiển thị hình ảnh sociogram của mạng lưới dựa trên các ước lượng trọng số các cạnh ω_{ij} , trong đó các dải xung quanh các cạnh tỷ lệ với độ rộng của các khoảng tin cậy 95%. Tại thời điểm này, mạng lưới ước lượng có thể được xử lý tương tự như một mạng lưới chuẩn của các ước lượng điểm, miễn là lưu ý bảo tồn sự không chắc chắn trong các phân tích tiếp theo. Để chứng minh sức mạnh của BISoN trong việc thực



2 (a) Hình ảnh hóa biểu đồ xã hội của một mạng được tạo ra từ các trọng số cạnh không chắc chắn. Chiều rộng của các cạnh biểu thị trọng số của chúng và các dải xung quanh các cạnh tỷ lệ thuận với chiều rộng của các khoảng tin cậy 95% của chúng. Các trọng số cạnh thấp được loại bỏ tùy ý khỏi biểu đồ này (mặc dù không phải là phân tích) cho mục đích hình ảnh hóa. (b) Phân phối sau của các trung tâm vector riêng cấp nút cho mỗi nút trong mạng. Các nút 1–4 là các nút được chỉ định cho phương pháp điều trị và các nút 5–8 cho đối chứng.

hiện các phân tích tiếp theo, nhóm tác giả đã thực hiện một phân tích hồi quy trong đó biến phản ứng là centrality eigenvector và biến dự đoán là một biến phân loại với hai loại, đại diện cho điều kiện điều trị. Để cải thiện mức độ phù hợp của mô hình, biến phản ứng đã được chuẩn hóa bằng cách trừ đi giá trị trung bình và chia cho độ lệch chuẩn. Cấu hình mô hình cho phân tích hồi quy là

$$\begin{aligned} EV_i &\sim \text{Normal}(\beta_{\text{Treat}[i]}, \sigma^2) \\ \beta_{\text{Treat}[i]} &\sim \text{Normal}(0, 1) \\ \sigma &\sim \text{Half-Normal}(0, 1) \end{aligned}$$

trong đó EV_i biểu thị độ tập trung vectơ riêng chuẩn hóa cho nút thứ i , $\beta_{\text{Treat}[i]}$ là tham số hiệu ứng tương ứng với điều kiện xử lý (kiểm soát hoặc xử lý) cho mẫu thứ i và σ là độ lệch chuẩn của phần dư. Mô hình hồi quy được điều chỉnh bằng bộ lấy mẫu Metropolis-Hastings tùy chỉnh của nhóm tác giả và sự hội tụ chuỗi được kiểm tra trực quan và với thống kê R. Các kiểm tra dự đoán sau được sử dụng để đảm bảo mô hình phù hợp với dữ liệu.

Phân phối sau của các tâm vectơ riêng không chuẩn hóa cho tám nút được thể hiện trong Hình 2b. Các tham số điều kiện xử lý có thể diễn giải là tâm vectơ riêng chuẩn hóa trung bình của các nút cho mỗi một trong hai điều kiện. Khoảng tin cậy 95% cho tham số kiểm soát β_{control} là $[-1,27, -0,36]$, với trung vị là $-0,83$ và cho tham số xử lý $\beta_{\text{treatment}}$ là $[0,35, 1,29]$, với trung vị là $0,85$. Một trong nhiều lợi ích của suy luận Bayesian là bản thân các phân phối sau của các tham số khác nhau có thể được sử dụng để có được những hiểu biết thống kê. Trong trường hợp này, câu hỏi thực sự quan tâm là sự khác biệt giữa hai điều kiện xử lý: $\beta_{\text{treatment}} - \beta_{\text{control}}$. Số lượng này có thể được tính toán đơn giản bằng cách trừ các hậu nghiệm tương ứng với tham số xử lý khỏi các hậu nghiệm tương ứng với tham số kiểm soát. Nhóm tác giả thấy rằng sự khác biệt về độ tập trung chuẩn hóa giữa các điều kiện xử lý, $\beta_{\text{treatment}} - \beta_{\text{control}}$, có khoảng tin cậy 95% là $[0,99, 2,33]$ và trung vị là $1,69$. Phân phối sau của sự khác biệt cũng được thể hiện trong Hình 2b. Điều này chỉ ra rằng điều kiện xử lý có vẻ tương quan mạnh với độ tập trung cao hơn.

Ví dụ ngắn gọn này chỉ phác thảo một số loại phân tích có thể được thực hiện trong khuôn khổ của nhóm tác giả.

V. THẢO LUẬN

Trong bài báo này, nhóm tác giả đã trình bày BISoN, một khuôn khổ chung, cực kỳ linh hoạt để mô hình hóa và phân tích các mạng xã hội. BISoN sử dụng các mô hình của các quy trình thu thập dữ liệu để định lượng sự không chắc chắn trong các trọng số cạnh. Các mô hình này rất linh hoạt và có thể được mở rộng để nắm bắt các hiệu ứng xã hội và phi xã hội bổ sung. Nhóm tác giả đã chứng minh cách sự không chắc chắn trong mạng sau đó có thể được truyền qua các phân tích tiếp theo, chẳng hạn như hồi quy liên quan đến tính trung tâm của mạng. Nhóm tác giả đã cung cấp mã ví dụ về cách triển khai khuôn khổ BISoN trong R với Stan tại: <https://doi.org/10.5281/zenodo.6603327>

BISoN thừa nhận sự không chắc chắn vốn có trong các trọng số cạnh suy ra, và bằng cách truyền bá điều này thông qua các phân tích hạ lưu có thể cung cấp các phân tích thống kê sắc thái, cân bằng và mạnh mẽ, tập trung vào việc định lượng các tác động sinh học và sự không chắc chắn xung quanh chúng. Cũng như truyền bá sự không chắc chắn vào

các phân tích hạ lưu, bản thân mô hình trọng số cạnh cũng có thể được điều chỉnh để đặt câu hỏi về sự thay đổi cá nhân trong tính bầy đàn, sự đáp lại của các hành vi có chỉ đạo và sự khác biệt cơ bản của các trọng số cạnh, trong số nhiều thứ khác. Nhóm tác giả coi BISoN là một tổng hợp và mở rộng tổng quát của nhiều phương pháp hiện có. Đặc biệt, đối với dữ liệu nhị phân, BISoN triển khai một biến thể của mô hình lỗi đo lường nhị thức beta, trước đây đã được sử dụng trong phân tích mạng xã hội động vật (Farine & Strandburg-Peshkin, 2015; Fink, 1997; Fuller, 1987). Cơ chế mô hình hóa nhiều hiệu ứng của BISoN được lấy cảm hứng từ các mô hình tuyến tính tổng quát và có thể được coi là phần mở rộng của các mô hình quan hệ xã hội (Kenny & La Voie, 1984; McCulloch & Searle, 2004). Khái niệm này được giới thiệu trong tài liệu về mạng xã hội động vật bởi (Whitehead & James, 2015) dưới dạng các chỉ số liên kết tổng quát. Khung BISoN có một số điểm tương đồng về mặt khái niệm với khung STRAND mới được phát triển gần đây, nhưng triển khai các mô hình đo lường được thiết kế riêng cho các loại dữ liệu thực nghiệm phổ biến và cho phép truyền bá sự không chắc chắn đến các phân tích hạ lưu. Một trong những lợi ích chính của khung của nhóm tác giả là tính linh hoạt vốn có của nó nhưng điều này cũng đưa ra một số cân nhắc khi sử dụng nó. Thiết lập các giá trị trước là một phần quan trọng của mô hình Bayesian và điều quan trọng là chúng phải phản ánh các kỳ vọng trước thực tế về các giá trị tham số (van de Schoot và cộng sự, 2021). Các giá trị trước được đưa vào mã ví dụ không có ý định là các giá trị mặc định tốt, thay vào đó, các giá trị trước phải được xác định bằng cách cân nhắc cẩn thận ý nghĩa của tham số, kỳ vọng trước về các giá trị có thể có của tham số và sử dụng các kiểm tra dự đoán trước (Conn và cộng sự, 2018). Điều quan trọng cần lưu ý là việc sử dụng các thuật ngữ bổ sung và thay đổi phần dự đoán của mô hình có thể thay đổi cách diễn giải các tham số mô hình, điều này có thể yêu cầu các giá trị trước cũng phải được thay đổi.

Một cân nhắc khác là các biện pháp trung tâm mạng dựa trên các cạnh nhị phân, chẳng hạn như bậc và tính bắc cầu, có cách giải thích hơi khác trong khuôn khổ này. Điều này là do, mặc dù một cặp cụ thể có thể chưa bao giờ được nhìn thấy tương tác trong, chẳng hạn như 100 mẫu, nhưng vẫn có khả năng chúng sẽ tương tác trong mẫu thứ 101. Các mô hình trọng số cạnh mà BISoN sử dụng thừa nhận điều này một cách tự nhiên, do đó có sự không chắc chắn xung quanh tất cả các trọng số cạnh. Nếu cần các trung tâm nhị phân, có hai cách có thể sử dụng chúng trong BISoN: (1) các trọng số cạnh có thể được ngưỡng hóa, để tạo ra phân phối nhị phân trên 0 và 1 cho mỗi cạnh; hoặc (2) các trọng số cạnh có thể được gán cho các danh mục và tư cách thành viên của một danh mục cụ thể (ví dụ: yếu hoặc mạnh) trở thành phép đo nhị phân mới, một lần nữa duy trì sự không chắc chắn đối với tư cách thành viên. Tùy chọn 1 có thể chỉ cần được áp dụng cho hậu nghiệm lấy mẫu bằng cách chọn điểm cắt có liên quan về mặt sinh học và áp dụng ngưỡng. Trong một số trường hợp, cách tiếp cận này có thể có vẻ thách thức vì nó đòi hỏi phải lập bản đồ kiến thức về miền và câu hỏi, trong ngữ cảnh, thành một điểm ngưỡng duy nhất để phân tích. Tùy chọn 2 yêu cầu gộp một phần trên các trọng số cạnh, giả sử rằng chúng được rút ra từ hỗn hợp các phân phối và gán mỗi trọng số cạnh cho từng loại kết nối với một xác suất nhất định. Cách tiếp cận này về mặt khái niệm tương tự như phương pháp phân loại liên kết xã hội do (Ellis và cộng sự, 2021) giới thiệu. Nếu về mặt sinh học, tin rằng có các loại kết nối khác nhau trong một

hệ thống là hợp lý, thì điều này có thể cung cấp một cách tự nhiên và có nguyên tắc để tính toán các phép đo mạng nhị phân.

Một cân nhắc cuối cùng về khuôn khổ mô hình của nhóm tác giả là, mặc dù các tác động phi xã hội có thể được mô hình hóa trong bộ dự báo, nhưng điều này không thể loại bỏ một số tác động phi xã hội và thay vào đó có thể hấp thụ tác động xã hội quan tâm, tùy thuộc vào hệ thống nhân quả cơ bản. Nhóm tác giả khuyến nghị rằng những tác động này phải được xử lý thận trọng và chỉ được sử dụng với các giả định rõ ràng về hệ thống nhân quả cơ bản.

Tính linh hoạt của khuôn khổ đề xuất của nhóm tác giả làm cho phương pháp tiếp cận mạnh mẽ, nhưng đi kèm với một số hạn chế. Mặc dù rất linh hoạt, một hạn chế quan trọng đối với BISoN là kiến thức về nỗ lực lấy mẫu sẽ được yêu cầu để mô hình hóa đúng các quy trình tạo ra các quan sát. Ở cấp độ thực tế, các mô hình BISoN thường sẽ cần phải phù hợp bằng cách sử dụng bộ lấy mẫu Markov chain Monte Carlo (MCMC). MCMC có thể chậm và tốn kém về mặt tính toán, đặc biệt là với các tập dữ liệu lớn. Nếu không bao gồm các hiệu ứng bổ sung, vấn đề này có thể được khắc phục một phần bằng cách tổng hợp dữ liệu ở cấp độ dyad và lắp các phiên bản thu gọn của mô hình (xem Thông tin hỗ trợ S8). Hơn nữa, các phương pháp MCMC không đảm bảo sự hội tụ, do đó cần phải kiểm tra thêm để đảm bảo rằng bộ lấy mẫu đã hoạt động chính xác (Cowles & Carlin, 1996; Draper, 2008). Những hạn chế này cùng nhau khiến quá trình xây dựng, lắp và kiểm tra mô hình trở thành một công việc phức tạp hơn là chỉ xây dựng mạng lưới từ các ước tính điểm. Tuy nhiên, để đổi lại những khoản đầu tư này, các nhà nghiên cứu sẽ có quyền truy cập vào một bộ công cụ mạnh mẽ có thể tận dụng tối đa dữ liệu khó kiểm được.

Khung BISoN là mở và có thể được mở rộng theo nhiều cách. Một khía cạnh đặc biệt hữu ích của phương pháp Bayesian là khả năng xử lý dữ liệu với sự không chắc chắn. Điều này cho phép dữ liệu bị thiếu được mô hình hóa trong mô hình thống kê mà không cần cần phải lắp lại các quy trình lắp mô hình. Hơn nữa, dữ liệu bị thiếu có thể được ước tính từ các biến khác, điều này có thể đặc biệt hữu ích trong các quần thể mà các đặc điểm như tuổi tác thường được ước tính hoặc không xác định (Little & Rubin, 2002). Một hướng tiềm năng khác cho khung là vượt ra ngoài các biện pháp xã hội tĩnh và thay vào đó mô hình hóa rõ ràng động lực của trọng số cạnh, như đã thảo luận bởi (Pinter-Wollman và cộng sự, 2014). Trọng số cạnh động có thể được mô hình hóa bằng cách kết hợp các mô hình chuỗi thời gian hồi quy tự động vào khung (Wei, 2013). Điều này sẽ giúp có thể mô hình hóa xu hướng chung trong trọng số cạnh, cách thay đổi trọng số cạnh lan truyền qua mạng và thậm chí mô hình hóa các quy trình luồng như bệnh tật hoặc truyền thông tin.

Mức đóng góp:

| | | |
|-------------------|----------|-------|
| Tên | MSSV | Tỉ lệ |
| Nguyễn Nhật Hoàng | 20520516 | 100% |

VI. TỔNG KẾT

Sự không chắc chắn ảnh hưởng đến tất cả các phân tích mạng xã hội, nhưng cho đến nay vẫn khó định lượng. Khung BISoN mà nhóm tác giả đã giới thiệu được thiết kế để giúp định lượng sự không chắc chắn và đảm bảo nó được bảo tồn thông qua các phân tích tiếp theo để đưa ra suy luận thống kê mạnh mẽ và đáng tin cậy. Khung của nhóm tác giả mang tính chung và có thể áp dụng rộng rãi cho hầu như mọi loại dữ liệu quan hệ xã hội. Trọng số cạnh ước tính từ các mô hình của nhóm tác giả có thể được sử dụng với nhiều loại phân tích hạ nguồn khác nhau trong khi truyền bá sự không chắc chắn. Nhóm tác giả tin rằng những ý tưởng được trình bày trong bài báo này chỉ là bước đầu tiên trong việc phát triển các mô hình thống kê mạnh mẽ, linh hoạt và mạnh mẽ cho các mạng xã hội. BISoN có tiềm năng cho phép nhóm tác giả đặt ra nhiều câu hỏi sắc thái và có được cái nhìn sâu sắc hơn về bản chất của xã hội.

REFERENCES

Tài liệu tham khảo trong báo cáo cũng là tài liệu tham khảo tương ứng trong bài báo.

[1] Croft, D. P., Darden, S. K., & Wey, T. W. (2016). Current directions in animal social networks. *Current Opinion in Behavioral Sciences, Behavioral Ecology*, 12(December), 52–58. <https://doi.org/10.1016/j.cobeha.2016.09.001>

[2] Whitehead, H. (2008a). *Analyzing animal societies*. University of Chicago Press.

[3] Cairns, S. J., & Schwager, S. J. (1987). A comparison of association indices. *Animal Behaviour*, 35(5), 1454–1469.

[4] Croft, D. P., James, R., & Krause, J. (2008). *Exploring animal social networks*. Princeton University Press.

[5] Weiss, M. N., Franks, D. W., Giles, D. A., Youngstrom, S., Wasser, S. K., Balcomb, K. C., Ellifrit, D. K., Domenici, P., Cant, M. A., Ellis, S., Nielsen, M. L. K., Grimes, C., & Croft, D. P. (2021). Age and sex influence social interactions, but not associations, within a killer whale pod. *Proceedings of the Royal Society B*, 288(1953), 20210617.

[6] Whitehead, H. (2008b). Precision and power in the analysis of social structure using associations. *Animal Behaviour*, 75(3), 1093–1099. <https://doi.org/10.1016/j.anbehav.2007.08.022>

[7] James, R., Croft, D. P., & Krause, J. (2009). Potential Banana skins in animal social network analysis. *Behavioral Ecology and Sociobiology*, 63(7), 989–997. <https://doi.org/10.1007/s00265-009-0742-5>

[8] McCulloch, C. E., & Searle, S. R. (2004). *Generalized, linear, and mixed models*. John Wiley & Sons.

[9] D. P. Kingma and M. Welling, “Auto-encoding variational Bayes,” 2013, arXiv:1312.6114. [Online]. Available: <https://arxiv.org/abs/1312.6114>

[10] S. Liu, “Wi-Fi Energy Detection Testbed (12MTC),” 2023, gitHub repository. [Online]. Available: <https://github.com/liustone99/Wi-Fi-Energy-Detection-Testbed-12MTC>

[11] “Treatment episode data set: discharges (TEDS-D): concatenated, 2006 to 2009.” U.S. Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, Office of Applied Studies, August, 2013, DOI:10.3886/ICPSR30122.