Annals of West University of Timisoara
Mathematics and Computer Science
60 (2024), 110–122

DOI: 10.2478/awutm-2024-0008

Investigating the Formation Phase of a New Online Social Network

Alin Brînduşescu

Abstract Over the last decade, we have witnessed the emergence of numerous online social media platforms and a growing research interest in understanding their underlying structures and mechanisms. This paper investigates the formation and evolution of the EcoNation social media platform, which is designed to foster collaboration for environmental sustainability.
The network structure of EcoNation is introduced and analyzed using a comprehensive dataset covering the platform's initial years of development. This analysis is conducted using the social network analysis toolbox, which enables the exploration of various aspects of user interactions, connectivity, and engagement patterns. By delving into the data, the aim is to uncover the dynamics that drive the growth and development of EcoNation's network.
The dataset includes information on user registrations, actions performed, validations, and connections formed over time. By examining this data, key trends and behaviors that have shaped the platform's evolution can be identified.
To facilitate further research and enable other scholars to build on this paper's findings, the exported data was made publicly available.

AMS Subject Classification (2020) 91D30; 05C82; 91C20
Keywords Social networks; centrality measures; network analysis; gamification; fraud-detection; simulation

1 Introduction

With over 5.07 billion users spending more than 2 hours a day online [6], social media has become an integral part of daily life for most people worldwide. While most users are concentrated on the top 10 social media platforms, new entrants emerge daily, aiming to solve a specific problem or address a particular interest group. Since all social media platforms are aiming more or less at the same population pool, new entrants need to find creative ways to attract users and keep them engaged. The quest for finding new ideas and technical solutions is challenging. From the first time the term "social media" was coined by the research community until 2020 there were published over 110.000 research articles related to "social media" [1]. One thing is certain: the social media platform should be continuously monitored and improved, while the emerging social network should be regularly consulted and evaluated.

The analysis of social network formation, growth, and characteristics has garnered significant research interest, leading to the discovery and validation of new patterns and laws that govern network structure and link creation. This research has illuminated various fundamental concepts that explain how social networks evolve and operate. Among the most well-known are the *small world phenomenon*, [3, 9, 11], *preferential attachment* [2, 13, 16], *scale free networks*.

The small world phenomenon describes how, despite large network sizes, most nodes can be reached from any other node through a small number of steps. The preferential attachment explains how nodes with more connections are more likely to receive new connections, leading to the emergence of hubs within the network. Scale-free networks, characterized by their power-law degree distribution, highlight how a few nodes (hubs) have significantly more connections than the average node, impacting the network's resilience and efficiency.

These concepts have been extensively validated by investigating data provided by online social networks. By analyzing large datasets from platforms like Facebook, Twitter, and LinkedIn, researchers have been able to observe these phenomena in real-world settings, thus confirming theoretical models and enhancing our understanding of social network dynamics.

The Fortunate Planet [14] started in 2020, during the COVID pandemic lockdown, as a movement to encourage people to recycle more and to reduce the waste disposed of in nature. In 2021, The Fortunate Planet launched EcoNation [15], a social media platform designed for users to collaborate on maximizing recycling and reducing the amount of waste disposed of in nature.

EcoNation is not a typical online social network, as users do not directly initiate connections with others they may know. The EcoNation underlying social network shares some similarities with academic social networks, the nodes of the networks are not necessarily connected because they know each other but by their trust in each other. In the case of the academic social network, the connection is created by citations, and in the case of EcoNation by validating the other node actions. Another differentiating factor is that the EcoNation social network is not ego-centering, and is not focused on the user but on the actions of the users. For the user, the social structure remains largely hidden; they can only observe the outcomes of other users' actions and their validations. While navigating through connections is not possible, the network structure plays an important role in the platform's development, growth, and maintenance.

To find its way in the challenging landscape of social media platforms EcoNation adopted the gamification strategy. Gamification as defined, since 2011, by Deterding et all "is the use of game design elements in non-game context" [7, 8]. The EcoNation platform incorporates several game design elements. Users are rewarded with points for their actions, and there are badges, leaderboards, and user levels. Challenges and goals are also featured, and the entire user experience was designed to resemble popular games like Pokemon GO. Many other platforms, such as StackOverflow and Reddit, also utilize gamification techniques. Their users build *digital social capital* by accumulating points for their contributions. EcoNation took it a step further by awarding users digital currency for each point earned. This currency can be used in an embedded marketplace where companies committed to a green future offer products or discounts in exchange for it.

2 Motivation

To sustain the platform's growth, it is crucial to identify the most relevant users. Recognizing these key contributors not only allows for effective incentivization to encourage their continued engagement but also provides a valuable resource for gathering feedback. Conducting surveys among these users can reveal insights into areas that need improvement, ensuring the platform evolves in ways that meet their needs and expectations. This dual approach of motivation and feedback collection helps maintain user satisfaction and drives the overall success of the platform. Additionally, having a clear picture of the emerging social network helps identify potential issues related to platform functionality. For example, if many users register from a specific region but do not perform any actions, it may indicate the need for region-specific adaptations.

Gamification alone is already enticing enough for hackers to exploit loopholes to boost their digital social capital. Introducing a digital currency that can be exchanged for physical goods amplifies this incentive, making the platform even more attractive to fraudulent behavior. The potential for real-world rewards significantly raises the stakes, necessitating robust security measures to protect the integrity of the system and ensure fair play among users.

The initial approach to mitigate fraudulent behavior involved introducing validation for user actions. However, this measure alone proved insufficient to fully address the problem. To strengthen the platform's defenses, a "trust capital" system was implemented, assigning a trust score to users based on their activity and reliability. Alongside this, the validation algorithms were significantly enhanced to better detect and prevent fraudulent activities [4].

The validation process introduced to mitigate fraudulent behavior has played a key role in shaping the emerging social network. This mechanism connects users through their actions and validation. Following its implementation, the monitoring of the network evolved from basic statistics, such as user sign-ups and actions performed, to a more sophisticated analysis using tools from social network science. These tools include components, centrality measures, reciprocity, shortest paths, and others, providing deeper insights into the network's dynamics and structure.

3 Network Model

The users and the actions they perform can be represented as sets, where U is the set of users and A is the set of actions. Using the sets a bipartite graph $G = (U, A, E)$ is constructed, where the vertices are divided into two disjoint sets, and the edges represent the relationships between users and actions. In this graph, the only possible type of relation is between a user and an action, formally defined as $E = \{(u, a)|u \in U, a \in A\}$.

The actions available in A have a limited number of types, constraints imposed by the design of the platform. We can define action types as $AT = \{at\ \textit{is a type action available on the platform}\}$. The validation is defined as an action and is part of the AT set. Among these action types, validation is considered an essential action and is included within the AT set.

To comprehensively describe the connection model, we introduce a set of functions that capture the relationships and interactions within the graph. These functions provide a mathematical framework to understand how users are connected through actions and validation.

- $ua : A \times U \to \{0, 1\}$ defined as:

$$ua(a, u) = \begin{cases} 1, & \text{if } u \text{ has performed action } a, \\ 0, & \text{otherwise.} \end{cases} \tag{3.1}$$

- $va : A \times U \to \{0, 1\}$ defined as:

$$va(a, u) = \begin{cases} 1, & \text{if } u \text{ has validated action } a, \\ 0, & \text{otherwise.} \end{cases} \tag{3.2}$$

- $valid : A \to \{0, 1\}$ defined as:

$$valid(a) = \begin{cases} 1, & \text{if } a \text{ has been validated}, \\ 0, & \text{otherwise.} \end{cases} \tag{3.3}$$

- $trusteduser : U \to \{0, 1\}$ defined as:

$$trusteduser(u) = \begin{cases} 1, & \text{if } u \text{ validated by the platform operator or his actions validated by min 3 trusted user} \\ 0, & \text{otherwise.} \end{cases} \tag{3.4}$$

- $connected : U \times U \to \{0, 1\}$ defined as:

$$con(u1, u2) = \begin{cases} 1, & \exists a \in A, va(a, u_1) = 1 \lor va(a, u_2) = 1, \\ 0, & otherwise. \end{cases} \tag{3.5}$$

The introduced functions provide a more accurate representation of how two users become connected within the network. In this model, the connection between users is not established through direct interaction, but rather through the validation process. Specifically, if one user performs one or more actions and another user validates at least one of these actions, a connection is formed between the two users. The function *connected* introduced by equation (3.5) captures very clearly this connection model.

3.1 Validation algorithm

During the initial phase, various algorithms were evaluated and implemented on the platform, leaving distinct traces on the network structure. Although not particularly relevant to this paper, one restriction applied by all validation algorithms was expected to have a significant impact: the proximity restriction.

Function `ValidateAction(`a, $u_{validation}$`):`
    $u_{action} = user\ with\ u(a, user) = 1$
    if $location(u_{validation})\ close\ to\ location(a)$ **then**
        if $not\ valid(a)\ and\ u_{action} \neq u_{calidation}$ **then**
            if $trusteduser(u_{validation})$ **then**
                $a = valid$
                if $u_{action}\ valid\ actions \geq 3$ **then**
                    $Add\ u_{action}\ to\ trusted\ users\ group$
                end
            end
        end
    end

Algorithm 1: Validation algorithm

In Figure 1 the action a_1 performed by user u_1 contains information about the location (latitude and longitude) where it occurred. For users u_2 and u_3 to be permitted to validate this action, they must be in close proximity to the location where the action was performed. It can be observed that user u_2 is close enough to the location where the action occurred, so the system will allow them to validate the action. However, user u_3 is too far from the location, and thus, the system will not permit their validation. In practice, the action location is recorded with high precision, as accurately as mobile devices permit. However, for the validation process, the proximity requirement is not as strict; the user validating the action only needs to be in the same city or area as the action location.

The validation proximity restriction is anticipated to lead to the formation of a social network characterized by numerous independent, well-connected components that are geographically localized. This means that the network will likely consist of distinct clusters, each corresponding to specific geographic areas, such as individual cities or countries. Each component will be tightly knit within its region, with strong internal connections but limited interaction with components outside its geographic boundary. As a result, the network will exhibit a clear spatial organization, where each major geographic area forms its cohesive network segment, reflecting the localized nature of the validation process.

4 Social Network Dataset

For the analysis, data on users and actions created on the EcoNation platform over two years was exported and anonymized. This involved several steps to ensure user privacy

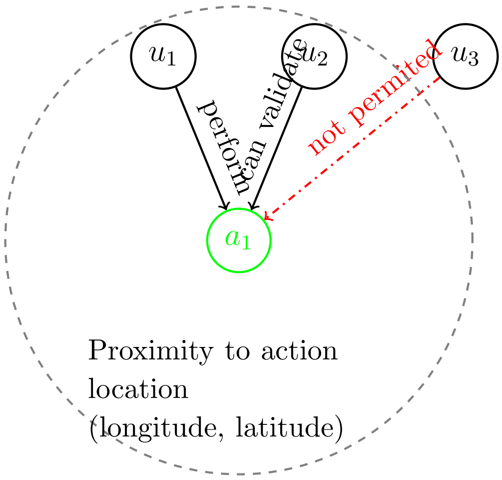


Figure 1: User Action Validation

and data protection. Firstly, all user identifiers were replaced with randomly generated identifiers, effectively decoupling any personal information from the actions performed on the platform. Secondly, the precise locations associated with actions were reduced to country codes, preventing any detailed geographic tracking. Thirdly, timestamps were clustered into 15 distinct intervals (T1 to T15), each covering the same duration, to obscure the exact times of user activities.

This anonymization process is crucial for several reasons. It safeguards the platform's integrity by protecting sensitive user data from potential breaches. Additionally, it ensures that the data handling practices are in full compliance with stringent data protection laws and regulations, such as GDPR. By anonymizing the data, we uphold users' privacy rights while still enabling meaningful analysis.

The anonymized dataset is now publicly available for download and analysis on the RecyclingSocialNetwork GitHub repository [5]. Researchers and analysts can access this dataset to conduct further studies, develop new insights, and contribute to the collective understanding of user behavior and network dynamics on the EcoNation platform.

The user collection consists of 2,481 individuals who registered on the EcoNation platform over the specified period. The distribution of these users across the 15 distinct time intervals is shown in Figure 2. The T3 interval, in particular, has the highest number of registered users. This spike is expected, as it coincides with a paid ad campaign that was run on both Facebook and Google, significantly boosting user acquisition during that period.

The distribution pattern changes notably in the later intervals, where there is a noticeable decline in the number of registered users. This decrease is associated with a strategic shift in the platform's user acquisition approach. Initially, EcoNation targeted a broad audience, aiming to attract random users to join the platform. However, during the later periods, the focus shifted towards attracting corporate users, particularly those with a vested interest in recycling and sustainability initiatives.

The comprehensive analysis of user registration data, reflected in the varied distribution across different intervals, provides valuable insights into the effectiveness of different marketing strategies and the evolving focus of the EcoNation platform. This data

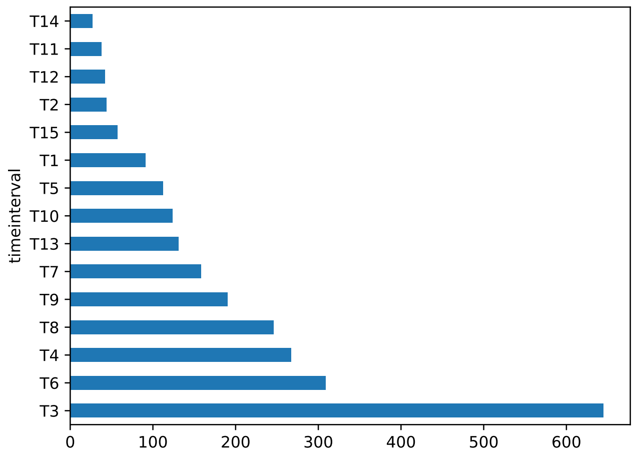
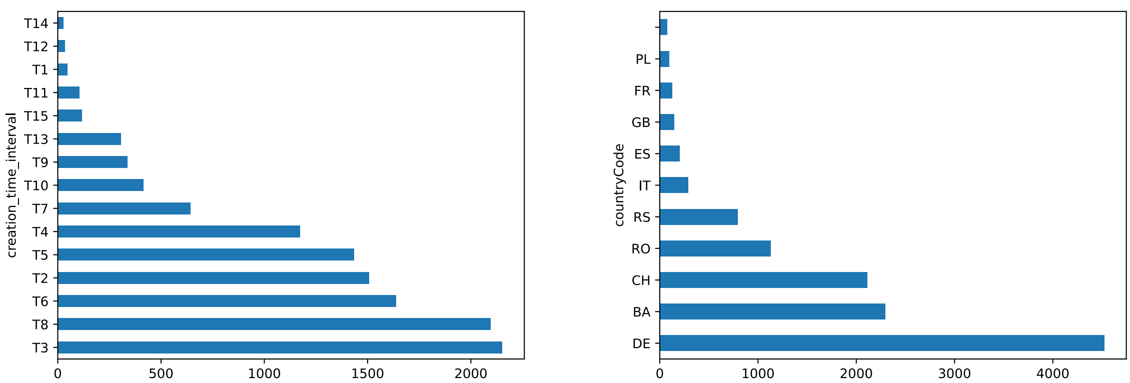


Figure 2: The distribution of users over the 15 intervals

is crucial for understanding user engagement patterns and for planning future growth strategies.

The actions collection consists of 12,035 actions and 11,405 verifications, leaving 630 actions unverified. The distribution of actions across time intervals and countries is displayed in Figure 3. Similar to the user dataset, there is a noticeable increase in actions during the T3 interval, which aligns with the paid ad campaign on Facebook and Google. Additionally, there was a significant rise in actions during the T8 interval, corresponding to a national cleaning and recycling campaign in Bosnia and Herzegovina and Serbia, which was supported by the EcoNation platform.



(a) Interval. (b) Country

Figure 3: The distribution of actions: a) over the 15 intervals, b) over the country

The country distribution reveals that most of the actions were performed in Germany, Bosnia and Herzegovina, Switzerland, Romania, and Serbia. These countries had the highest activity levels due to the platform running more ad campaigns and participating in national cleaning and recycling projects within these regions.

5 Findings

The number of registered users on the platform significantly exceeds the number of users who actively contributed actions. While over 2,400 users have registered, only 561 of them have engaged with the platform by performing actions or validations. This discrepancy highlights a notable gap between user registration and active participation.

In the initial time interval, this disparity is evident with a modest number of actions recorded - only 47 actions from 91 registered users. The number of validations is even lower, totaling just 5, which translates to an engagement rate of only 5.5% among users. This limited validation activity can be attributed to the fact that the validation algorithm was introduced partway through this period and was not integrated from the start. As a result, the system's ability to validate actions was not fully operational at the beginning, which impacted the number of validations performed.

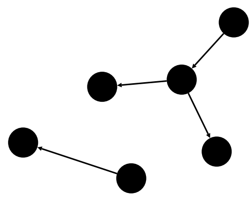


Figure 4: The EcoNation Network after first-time interval

Figure 4 illustrates the network structure at the end of this first interval. The network visualization provides a snapshot of how the limited number of actions and validations influenced the network's development and connectivity during this early stage.

In the subsequent time intervals, improvements in platform usability and targeted ad campaigns significantly fueled network growth and boosted user engagement. During the T3 interval, the platform saw the addition of 645 new users who collectively performed 2,151 actions and made 2,748 validations, resulting in the creation of 136 new connections. Figure 5 illustrates the structure of these connections established during the T3 interval.

In summary, the final network, as depicted in Figure 6, comprises 561 users and 11,405 connections. This extensive network illustrates a significant degree of growth and engagement over time. The network structure reveals the emergence of well-connected clusters, which are closely associated with the geographic locations where the actions were performed. These clusters reflect the spatial distribution of user activities and highlight how actions in specific regions contribute to the formation of dense, interconnected groups within the network. The visualization not only shows the overall connectivity but also underscores the influence of location-based actions on the network's structure.

Surprisingly, the final network comprises a large component that covers 88.8% of the network, along with several very small components consisting of 2 to 5 users each (see Table 1). This outcome is unexpected given the proximity constraint introduced in the validation algorithm (see Algorithm 1). The anticipation was that the validation proxim-

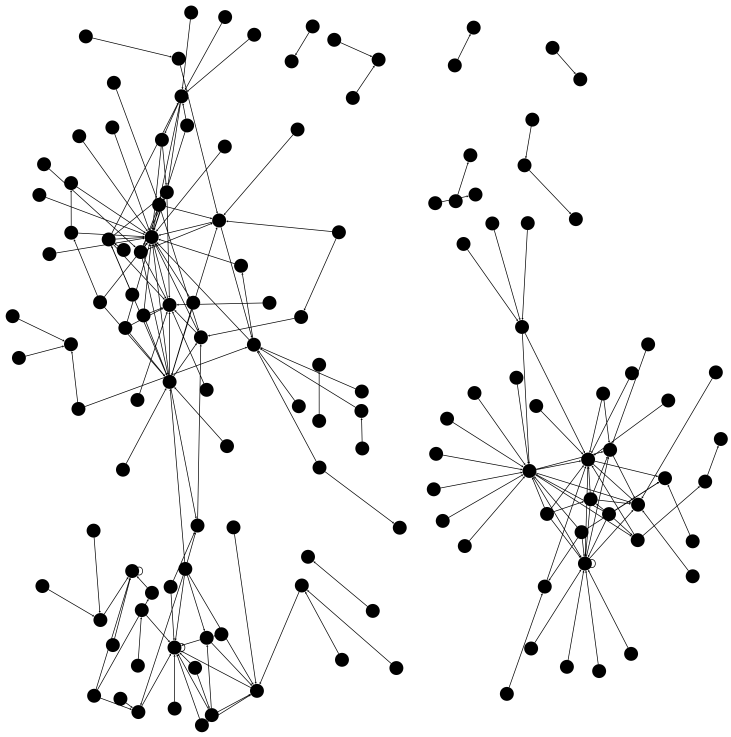


Figure 5: Connections created over a time interval (T3)

Table 1: Final Network components

Number of Nodes	Number of components	Percentage
498	1	88.8%
5	1	0.9%
4	2	1.4%
3	2	1%
2	22	3.9%

ity constraint would result in geographically close countries forming connected clusters, but not that the majority of the network would merge into a single, large component.

This large component indicates a higher-than-expected level of interconnectivity among users across different regions, despite the geographic limitations set by the validation process. It suggests that users in different locations found ways to bridge connections, potentially through overlapping validation activities or travel patterns.

With an average path length of 4.5 links within the largest component, it is safe to state that the EcoNation social network exhibits the characteristics of a small world. This average path length is shorter than the classic "six degrees of separation" concept [12], but slightly longer than the average path length of 3.57 calculated for Facebook users [9].

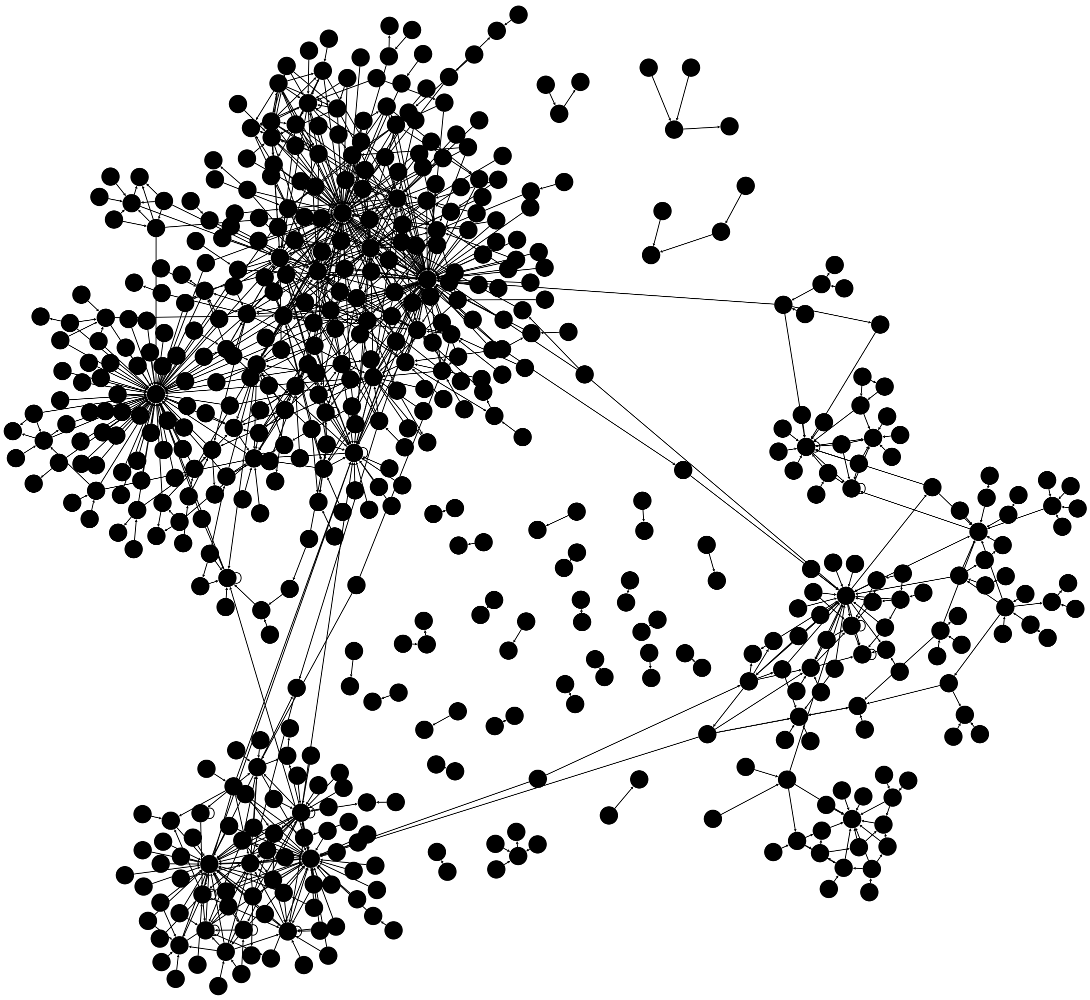


Figure 6: Network at the end of the last interval

5.1 Preferential attachment

Another unexpected phenomenon revealed by the analysis of the EcoNation social network is "preferential attachment" [2]. Given the platform's design, the expectation was for a very well-distributed network of connections among users. However, the analysis showed that certain users accumulated significantly more connections.

Table 2: Final Network User Degree Centrality

Degree Centrality	Number of users	Percentage
greater than 70	3	0.06%
between 30 and 40	5	0.1%
between 20 and 29	5	0.1%
smaller than 20	485	97.4%

The data displayed in Table 2 displays the distribution of users based on the degree centrality measures. The degree centrality (5.1) is the simplest centrality measure but it is the one capturing the best preferential attachment.

$$C_{Degree}(u) = \sum_{u_i \in U} con(u, u_i) \tag{5.1}$$

It can be observed that only 0.06% of the users have a degree centrality greater than 70, which is more than three times higher than that of the majority of users. This phenomenon cannot be explained by the classical "preferential attachment" theory, as users are not directly choosing whom to connect with. Instead, it relates to the "nature" of certain individuals.

Some people are more inclined to contribute and have a stronger desire to get involved in the community, leading to a higher degree of centrality. These highly engaged users tend to perform more actions and validations, thereby becoming central nodes within the network. Their active participation not only increases their own connectivity but also fosters greater interaction and connectivity among other users.

This behavior underscores the importance of individual user contributions in shaping the overall network structure and confirms that the network is scale-free. Identifying these highly connected users is crucial for the platform, as it can help foster further growth and address potential issues associated with scale-free networks. By recognizing and supporting these key influencers, the platform can enhance user engagement and mitigate problems such as network vulnerability.

5.2 Reciprocation

One network mechanism presenting an interest for the platform operator is the *reciprocation: "the tendency of vertex pairs to form mutual connections between each other"* [10].

In EcoNation, the validation process is directional: the user performing the validation creates a directed edge from themselves to the user who carried out the action ($u_{validation} \rightarrow u_{action}$). This directional validation is a built-in feature of the platform, and it is expected that some users will cross-validate each other's actions. However, if a pair of users engage in a continuous cycle of creating and validating actions within a closed loop, it could indicate a natural occurrence in regions with few users. Nonetheless, due to the platform's reward system, this behavior could also signal potential fraudulent activity.

To address this, it is important to monitor the system for any sudden increases in reciprocation values, especially within small and newly established components. Such anomalies should trigger alerts for further investigation. Manual analysis of these components is necessary to ensure that the observed patterns of reciprocation are legitimate and not indicative of manipulative behavior aimed at exploiting the reward system.

The final network's reciprocation rate is 0.2241, showing that most users do not reciprocate by validating the actions of those who validate their actions. This value is within expected parameters because the network structure is hidden from users, making it difficult for them to know whom to reciprocate with—unless they are coordinating their actions, which could indicate potential fraudulent behavior. Monitoring newly created groups for an increased reciprocation rate is a key strategy in detecting potential fraudulent activity. When an unusually high rate of reciprocation is observed, it serves as a red flag that prompts the platform to initiate manual verification. This process acts as the first level of fraud detection, ensuring that suspicious behavior is identified and investigated early on, helping to maintain the integrity of the platform and prevent the escalation of fraudulent activities.

6 Conclusion

Our investigation highlights several characteristics of the EcoNation social network and reveals some surprising findings. Despite the built-in constraints, the underlying network displays both *scale free* and *small world* patterns. Additionally, although *preferential attachment* was anticipated to be much smaller, it is still expected that in the future will diminish as the user base expands, but for now, it remains prominent. Notably, due to the high level of contribution from certain users, these individuals are forming connections with approximately 15% of the network.

References

[1] T. Aichner, M. Grünfelder, O. Maurer, D. Jegeni, *Twenty-five years of social media: a review of social media applications and definitions from 1994 to 2019*, Cyberpsychology, Behavior, and Social Networking **24 (4)** (2021), 215-222.

[2] A.-L. Barabási, R. Albert, *Emergence of scaling in random networks*, Science **286 (5439)** (1999), 509-512.

[3] G. Blank, C. Lutz, *The social structuration of six major social media platforms in the united kingdom: Facebook, linkedin, twitter, instagram, google+ and pinterest*, in: Proceedings of the 7th 2016 International Conference on Social Media & Society, 2016, 1-10.

[4] A. Brînduşescu, V. Monescu, N. Minculete, *Simulation of trust propagation in community validated gamification*, in: Proceedings of the VII International Scientific Conference Mathematical Modeling, 2023, 52–55.

[5] A. Brînduşescu, *Recycling social network dataset*, RecyclingSocialNetwork, https://github.com/alinoz77/, 2024.

[6] D. Chaffey, *Global social media statistics research summary may 2024*, https://www.smartinsights.com/, 2024. [Online; accessed 1-June-2024].

[7] S. Deterding, D. Dixon, R. Khaled, L. Nacke, *From game design elements to gamefulness: defining "gamification"*, in: Proceedings of the 15th International Academic MindTrek Conference: Envisioning future media environments, 2011, 9–15.

To foster collaboration and enable other researchers to analyze the data the complete dataset exported from EcoNation is available for download for free from *https://github.com/alinoz77/RecyclingSocialNetwork/* [5]. A multitude of statistical and social network analyses are possible and we hope we can find new direction from the future research publications.

[8] S. Deterding, R. Khaled, L. Nacke, D. Dixon, et al, *Gamification: Toward a definition*, in: CHI 2011 Gamification Workshop Proceedings, volume 12, 2011, 1–79.

[9] S. Edunov, C. Diuk, I. Onur Filiz, S. Bhagat, M. Burke, *Three and a half degrees of separation*, Research at Facebook, 694, 2016.

[10] D. Garlaschelli, M. I. Loffredo, *Patterns of link reciprocity in directed networks*, Physical Review Letters **93 (26)** (2004), 268701.

[11] S. Kemp, *Essential facebook statistics and trends for 2023*, Report DATAREPORTAL, https://datareportal.com/essentialfacebook-stats, 2023.

[12] S. Milgram, *The small world problem*, Psychology Today **2 (1)** (1967), 60–67.

[13] A. Mislove, H. S. Koppula, K. P. Gummadi, P. Druschel, B. Bhattacharjee, *Growth of the flickr social network*, in: Proceedings of the First Workshop on Online Social Networks, 2008, 25–30.

[14] The Fortunate Planet AG, *The fortunate planet*, https://thefortunateplanet.com/, 2022. [Online; accessed 1-June-2024].

[15] The Fortunate Planet AG, *Econation*, https://econation.me/, 2024. [Online; accessed 1-June-2024].

[16] M. Wang, G. Yu, D. Yu, *Measuring the preferential attachment mechanism in citation networks*, Physica A: Statistical Mechanics and its Applications **387 (18)** (2008), 4692–4698.

Alin Brînduşescu

West University of Timişoara, Bd. V. Pârvan 4 300223 Timişoara, Romania

The Fortunate Planet AG, Baar, Switzerland

E-mail: `alin.brindusescu@e-uvt.ro, alin@thefortunateplanet.com`

Received: 3.08.2024

Revised: 12.08.2024

Accepted: 24.08.2024