Annals of GIS

ISSN: 1947-5683 (Print) 1947-5691 (Online) Journal homepage: www.tandfonline.com/journals/tagi20

Exploring social interaction patterns and drivers in VGI communities using a custom geovisual analytics tool

Guiming Zhang, Wei Luo, Mingguang Wu & Liangfei Ye

To cite this article: Guiming Zhang, Wei Luo, Mingguang Wu & Liangfei Ye (2025) Exploring social interaction patterns and drivers in VGI communities using a custom geovisual analytics tool, Annals of GIS, 31:3, 413-431, DOI: 10.1080/19475683.2025.2497026

To link to this article: https://doi.org/10.1080/19475683.2025.2497026

ANNALS OF GIS
2025, VOL. 31, NO. 3, 413–431
https://doi.org/10.1080/19475683.2025.2497026

Taylor & Francis
Taylor & Francis Group

Exploring social interaction patterns and drivers in VGI communities using a custom geovisual analytics tool

Guiming Zhang[a], Wei Luo[b], Mingguang Wu[c] and Liangfei Ye[d]

[a]Department of Geography & the Environment, University of Denver, Denver, CO, USA; [b]GeoSpatialX Lab, Department of Geography, National University of Singapore, Singapore; [c]College of Geographic Sciences, Nanjing Normal University, Nanjing, China; [d]Department of Sociology, Texas A&M University, College Station, TX, USA

ABSTRACT
Social interactions among data contributors are essential to the success of many VGI (volunteered geographic information) projects, and the patterns of such interactions are often shaped by geographies in which the contributors are situated. However, there is a lack of investigations on geographic context's influences on VGI contributor behaviour and interaction. This study explores patterns and drivers of inter-contributor species identification activities in the iNaturalist biodiversity citizen science community using a custom geovisual analytics tool integrating visualization and analysis of social networks in geographic context. Geovisual explorations of the iNaturalist social network revealed that the frequency and intensity of species identification interactions in iNaturalist are influenced by three geographic contextual factors, namely, the geographic distance, land cover composition similarity and species taxon composition similarity between observer contributors and identifier contributors. The findings align with social theories concerning the key forces driving the formation of social interactions in a social network, wherein geographic distance reflects physical proximity and land cover composition similarity, and species taxon composition similarity reflects homophily effects. The geovisual analytics tool effectively facilitates exploring patterns and drivers of social interactions and offers a new lens through which to examine the social and geographic dynamics of social interactions in VGI communities.

ARTICLE HISTORY
Received 10 October 2024
Accepted 17 April 2025

KEYWORDS
Volunteered geographic information (VGI); spatial social network; interaction patterns and drivers; geovisual analytics; iNaturalist

1. Introduction

Over the past 20 years or so, the phenomenon of volunteered geographic information (VGI) has emerged and revolutionized the way geographic data is generated, shared, and utilized (Elwood, Goodchild, and Sui 2012; Goodchild 2007; Yan et al. 2020; G. Zhang 2021). The prosperity of VGI is facilitated by not only the technological enablers but also the countless volunteers who actively observe natural and built environments and collaboratively contribute to VGI (Huang et al. 2024). Volunteers have created collaborative communities like OpenStreetMap and iNaturalist, which exemplify the power of collective intelligence of the public in mapping the world and documenting Earth's biodiversity (Haklay and Weber 2008; Unger et al. 2021).

The success of VGI projects, in many cases, hinges on the intricate social interactions among members of the VGI community. These interactions represent a complex web of communication, collaboration and learning in the processes of co-creating VGI data. For instance, OpenStreetMap's success can be largely attributed to its vibrant community that actively engages in discussions, resolves conflicts, and collectively improves the map's accuracy and detail through co-editing activities (Mooney and Corcoran 2014). Biodiversity citizen science as a form of VGI has become pivotal in the collection of Earth's biodiversity data (Haklay 2021; Huang et al. 2024; Zhang 2021). Projects such as eBird and iNaturalist harness the enthusiasm and efforts of citizens to document species distributions, contributing to large-scale databases crucial for biodiversity research and conservation (Mesaglio and Callaghan 2021; Wood et al. 2011). The social interactions within these communities are fundamental to their functionality. In iNaturalist, community members submit species observations, engage in discussions to share experiences, and assist with species identification through vetting and/or voting on species identifications. Such social interactions not only create a supportive environment that encourages learning and participation but also improve the quality of the resulting species observations dataset (Aristeidou et al. 2021; Unger et al. 2021). The success of these VGI projects is deeply rooted in the social interactions among members of the collaborative communities.

CONTACT Guiming Zhang guiming.zhang@du.edu Department of Geography & the Environment, University of Denver, Denver, CO, USA

Therefore, analysing social interactions within VGI communities to discover interaction patterns and their drivers is essential for understanding the dynamics of VGI communities. Such analyses can address important questions, such as which VGI community members tend to interact more frequently with other members, where these interacting members are situated, and what demographic, socioeconomic and geographic factors influence the strength of their interactions. Social interactions within VGI communities can be modelled as spatial social networks in which nodes represent community members (actors) linked to specific places (e.g. a person's home city) and edges represent social interactions among nodes (e.g. exchange or flow of information) (Andris 2016; McCulloh, Armstrong, and Johnson 2013; Ye and Andris 2021). In the analysis of these networks, it is crucial to consider geographic context because actor's geographic location often implies the geography that may affect their behaviour and decision-making in interacting with others (Andris 2016). Analysing the networks provides valuable insights into the patterns of collaboration among community members and sheds light upon the drivers (the 'why' factors) that may have shaped the interaction patterns (Ye and Andris 2021; G. Zhang, Gong, and Zhu 2024a). Such analyses can facilitate the understanding of the social and geographic dimensions of interactions within VGI communities. This dual perspective is crucial for grasping the full scope of interactions and their implications for VGI project success.

Traditional social network analysis tools focus on exploring the social structures within communities (Oliveira and Gama 2012), but lack the capability to handle spatial data or integrate geographic context into the analysis. Thus, they typically do not consider the spatial dimension that influences social ties and information exchange, leading to a disconnect between social and geographic analyses (Luo and MacEachren 2014). Efforts have been made in integrating geographic context into spatial social network analysis through two closely related means: geovisual explorations and spatially informed analyses. Geovisual explorations, by visualizing and interactively exploring social networks situated in geographic spaces (e.g. on maps), aid in discovering interaction patterns in the network and formulating geographic theory-informed hypotheses to explain such patterns (Andrienko and Andrienko 2006). In spatially informed analyses, patterns visually extracted during geovisual explorations are quantified using spatial metrics (Luo 2016; Luo et al. 2014), hypotheses are tested, and findings are interpreted in the context of existing or new social and geographic theories (Andris et al. 2021; Radil, Flint, and Tita 2010; G. Zhang, Gong, and Zhu 2024a). Existing studies with the 'visualization' focus of spatial social networks have developed a variety of geovisualization interfaces to visualize geospatial networks such as social networks, trade and migration, and traffic and transport networks (Schöttler et al. 2021). Most network visualizations are superimposed on background maps representing the network's geographic context, but only some support user interactivity (Schöttler et al. 2021), which is essential for supporting effective interactive exploration of the data for pattern discovery and hypothesis formulation. Furthermore, geovisualization and geovisual analytics tools often need to be tailored to specific problem contexts to be most useful, although their design principles are transferable (Jin, Endert, and Andris 2024; Robinson 2017). There is hardly a generic geovisualization tool that suits all problems. For instance, the geovisual analytics interface for analysing international trade (Luo et al. 2014) obviously requires a different set of visual elements and interactivities than the one for examining family structures (Giordano, Cole, and Le Noc 2022) or for exploring knowledge graph (Wang, Li, and Gu 2023). Regarding VGI, most geovisual analytics tools are developed to investigate the data contribution patterns of VGI contributors at the individual level or group level through the lenses of spatial, temporal, or thematic characteristics. Examples include Pascal Neis's suite of tools (https://resultmaps.neis-one.org/.) and Crowd Lens (Quinn and MacEachren 2018) for exploring OpenStreetMap data. Few tools offer means for examining social interactions among VGI contributors. One exception is the geovisualization framework for conducting exploratory analysis of VGI contributor's participation patterns, which provides an interactive social network visualization (G. Zhang 2024). However, the network visualization is not situated in geographic space, nor does it provide network analysis functionalities.

Among studies with the 'analysis' emphasis of spatial social networks, some develop new or extend existing network metrics to reflect the spatial properties of the network (Gao et al. 2018; Sarkar et al. 2019). Others 'spatialize' the network in geographic space to examine, for example, how actor's relative location in geographic space may affect their structural position and interaction in the network (Andris et al. 2021; Radil, Flint, and Tita 2010). These analyses primarily consider geographic proximity (e.g. distance or spatial adjacency) as an indicator of geographic context. On analysing VGI-related social networks, most studies use social media VGI (Feng, Huang, and Sester 2022) to understand social media user interaction patterns, for instance, during disaster events (Xu and Qiang 2022) and public health crises (Gong et al. 2023). Many studies have found a general 'distance-decay' effect (Kottwitz, Zhang, and Xu 2023) that social media users tend to interact more with other users who are geographically closer. Others examine how geographic contextual factors such as demographics and political affiliation may impact social interactions in the network (Gong and Ye 2021). As for non-social media VGI, studies often adopt traditional social network analysis methods to elucidate patterns of interactions (e.g. co-editing activities in OpenStreetMap), but short of explicitly accounting for actor's geographic context to understand the spatial dimensions of the 'why' factors (Jusup et al. 2022; Mooney and Corcoran 2014; Sarkar and Anderson 2022; Sbrocchi et al. 2022; Zhang et al. 2021). Among existing spatial social network analyses attempting to integrate geographic context, geographic proximity still appears to be the dominant indicator. While spatial proximity is a good first approximation, geographic context should encompass any relevant aspects of the physical and social environments that may influence actor's behaviours and interactions (Andris 2016; Emch et al. 2012).

In general, geovisual exploration and spatially informed analysis of social networks are largely problem domain-specific and are often not integrated into a single system. This separation can create a disconnect between geographic context-aware visualization and the analysis of social networks. One notable exception is the recent SNoMaN visual analytic tool (Jin, Endert, and Andris 2024). The tool, offering several types of visual diagrams for spatial social network mapping and analysis, appears to be a more generic and unified framework that visualizes and analyzes spatial social networks from multiple domains (e.g. Mafia members network, global flight network, and food-donation-sharing network).

Examining social interactions within VGI communities requires integrating the multifaceted geographic context into the visualization and analysis of social networks. Here, geographic context encompasses more than just the node's location; it includes any geographic contextual information relevant to the node. For example, species identification interactions in the iNaturalist VGI community can be modelled as a social network, where contributors (observers and/or identifiers) are represented as nodes and observer-identifier species identification interactions are represented as directed, weighted edges (with the number of identifications between observer and identifier as the weight). Each node has the centroid of the observer's observation locations as its geographic location and is associated with additional attributes, such as the land cover frequency distribution across the observation locations (G. Zhang, Gong, and Zhu 2024b). Geovisual explorations of such a spatial social network require a custom geovisual analytics tool, as existing tools do not seem to fully support these needs. For example, the SnoMaN tool currently only supports unweighted, undirected networks and does not accommodate node attributes beyond geographic coordinates (Jin, Endert, and Andris 2024).

This study explores patterns and potential drivers of social interactions in VGI communities, using the iNaturalist collaborative biodiversity observation VGI community as a case study. It explores research questions such as: which iNaturalist contributors interact more frequently with other members of the community? What geographic and non-geographic factors influence these interactions? Are there distinct subgroups within the network, and where are these subgroups located? Are the members of the subgroups geographically clustered? To answer such questions, a spatial social network was constructed based on inter-contributor species identification activities in iNaturalist. Additionally, a custom geovisual analytics tool was developed to visualize and analyse the network in geographic context. This approach enables the discovery of interaction patterns and the exploration of how such patterns may be associated with geographic contextual factors, such as location and environmental characteristics (e.g. land cover). This study offers a novel perspective on the social and geographic dynamics of social interactions in VGI communities.

iNaturalist is a global online community for nature enthusiasts to report and identify biodiversity observations. The choice of species that contributors observe and/or identify often depends on their personal taxonomic interest and expertise. Contributors from the same geographic region are more likely to focus on similar species within that region, and the level of enthusiasm and expertise in reporting and/or identifying species can vary drastically across contributors, ranging from novices to experts. iNaturalist also functions as a social media platform where community members can follow each other and form topic groups (e.g. BioBlitz projects). Social theories suggest several key drivers of interactions in a social network, including homophily, reciprocity, proximity, prestige, social conformity, transitivity, and balance (McCulloh, Armstrong, and Johnson 2013). Among these drivers, the most relevant to inter-contributor species identification interactions in the iNaturalist community are expected to be homophily (social actors sharing common interests and beliefs are more likely to establish social interactions), proximity (social actors who have closer organizational or physical distance are more likely to interact), prestige (important actors hold great influence within the social network and tend to form more social links), reciprocity (social actors tend to reciprocate direct interactions initiated by others), and transitivity (social connections are likely to form through mutual connections with a third party). Particular attention was given to the effects of geography proximity and homophily in shaping social interactions in iNaturalist, which also guided the development of the custom geovisual analytics tool.

The organization of the remainder of this article is as follows. Section 2 introduces the datasets used in this study. Section 3 details the iNaturalist social network modelled from the datasets. Section 4 presents the custom geovisual analytics tool. Section 5 employs the geovisual analytics tool to explore social interaction patterns in iNaturalist and their potential associations with geographic contextual factors. Section 6 ends the article with a discussion and conclusions.

2. Datasets

2.1. iNaturalist data

This study used VGI data from iNaturalist, a biodiversity-themed citizen science community for citizen scientists to contribute and share species observations from around the world (Unger et al. 2021). Each species observation submitted to iNaturalist (e.g. species photos and/or audios) is assigned a unique record id and has information regarding observation location (latitude and longitude), observer (user id and login) and observation time. An initial species identification can be entered by the observer. Other members of the iNaturalist community can review the submission and propose identifications that may confirm or contradict the initial identification. A majority-vote approach is then adopted to determine a final species identification for the observation. Such observations constitute 'raw' observations. Observations with vetted identifications and meeting additional quality control requirements (e.g. non-captive organism and with geographic location) are marked 'research-grade' (iNaturalist 2024) and only 'research-grade' observations are published through the Global Biodiversity Information Facility. Information about the identifier (user login) who first proposed the vetted identification and the time are appended to 'research-grade' observations only.

This study focuses on examining inter-contributor species identification interactions in iNaturalist using a geovisual analytics tool. 'Research-grade' observations thus are essential for reconstructing a social network representing such social interactions (see section 3 for details). Nonetheless, 'raw' observations are also necessary for deriving contributor attributes, such as the approximate geographic location of a contributor and the environment characteristics at observation locations. 'Research-grade' observations were downloaded from the Global Biodiversity Information Facility (Ueda 2022) (~55 million records as of 31 December 2022). 'raw' observations were exported from the iNaturalist website (iNaturalist 2022) (~139 million records as of 31 December 2022). iNaturalist data in the full year of 2022 were used in this study, including 'raw' observations made in 2022 (~36 million records) and 'research-grade' observations identified in 2022 (~16 million records) regardless of when the observations were made.

2.2. Land cover data

This study intends to develop a geovisual analytics tool that integrates geographic context into exploring social interaction patterns in iNaturalist (and potentially other VGI communities). Here land cover was adopted to characterize the physical environments wherein contributors conduct observations. Land cover is most relevant considering iNaturalist contributors observe and/or identify species of all taxa, whose habitats can be broadly represented by land cover. The yearly (2001–2019) MODIS global land cover type dataset (MCD12Q1 Version 6; 500 m resolution) (Sulla-Menashe and Friedl 2018) was used to extract the land cover type at iNaturalist observation locations (the most recent 2019 land cover data were used to extract land cover type for observation locations after 2019).

3. iNaturalist social network

Different types of social networks have been modelled from iNaturalist data to analyse user (contributor) behaviour and interactions. Liu and Jiralerspong (2023) used iNaturalist 'research-grade' observations from a small region in Canada to build a user–user network, where each distinct user is represented as a node, and an undirected, unweighted edge connects two users if they have identified at least one common species. The network was then analysed to uncover users with similar taxonomic interests in their identification activities. Tupikina et al. (2021) used iNaturalist data from 11 City Nature Challenge events organized in three cities during 2017–2020 to model a social network. In their approach, each user is represented as a node, and an edge links two users if one has ever identified species observations submitted by the other. The undirected edge is weighted by the number of identification interactions. In this study, inter-contributor species identification interactions within the global iNaturalist community in

2022 were also modelled as a social network. In this network, a node is represented as an iNaturalist contributor (i.e. an observer or identifier), and an edge is a directed link from an observer (who submit observations) to an identifier (who identifies observations and can also be an observer). Each edge is weighted by the number of species identified between the two contributors (G. Zhang, Gong, and Zhu 2024b).

Nodes and edges also have associated attributes that are derived from the datasets (section 2) to represent geographic context (Figure 1). Each node (contributor) has three attributes: geographic location (approximated with the centroid of the contributor's observation locations; e.g. home city), land cover composition (the relative frequency distribution of observation locations over land cover type categories), and species taxon composition (the relative frequency distribution of species observations over species taxon categories at the kingdom level). For an observer, the composition was computed based on observations submitted by the observer (referred to as observer land cover/species taxon composition). For an identifier, the composition was computed based on observations identified by the identifier (referred to as identifier land cover/species taxon composition). If a node assumes both observer and identifier roles in the network, it would be associated with both observer and identifier land cover/species taxon compositions. Based on these node attributes, three corresponding edge attributes can be derived: geographic distance (i.e. the great circle distance between observer geographic location and identifier geographic location), land cover composition similarity (the cosine similarity between observer land cover composition and identifier land cover composition), and species taxon composition similarity (the cosine similarity between observer species taxon composition and identifier species taxon composition).

Among the node attributes, species taxon composition represents a contributor's personal interest in observing a species, and species taxon composition similarity indicates to what extent that interest is shared with other contributors in the community. Geographic location and land cover composition imply geographic proximity and characteristics of the environment in which species observations are conducted, respectively. They reflect the geographic context of the iNaturalist social network.

The full iNaturalist network consisted of 234,126 nodes and 1,063,274 edges. Visualizing and analysing such a large network poses significant computational challenges that exceed the capabilities of any web-based geovisual analytics tool (Jin, Endert, and Andris 2024). To reduce the size of the network, edges with weight below a threshold value were removed. While this filtering process excludes edges representing a small number of identifications between observers and identifiers, and thus degrades the fidelity of the reduced network to the original, many of these edges can be considered as 'noise' interactions. For example, they may represent ad-hoc identifications provided by the 'one-off' type contributors who submitted only a few observations or identifications before leaving the community (Zhang, Gong, and Zhu 2024b). After experimentation, an edge weight threshold of 10 was adopted, as it provided a reasonable balance between maintaining fidelity to the original network and reducing computational load. This filtering reduced the network to 8314 nodes and 17,000 edges. Furthermore, this study focuses on visualizing and analysing the largest connected component of this reduced network, referred to as the

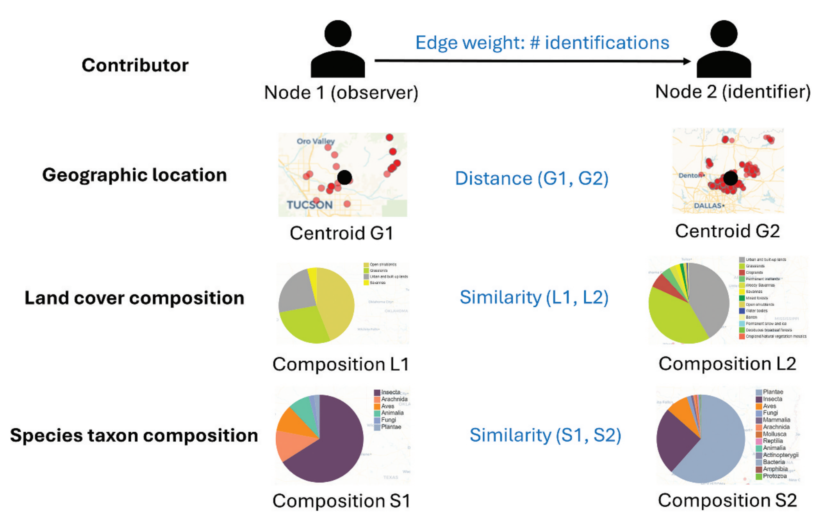

Figure 1. Node attributes and edge attributes in the iNaturalist social network illustrated with an observer-identifier pair (i.e. an edge).

iNaturalist network thereafter. This component contains 94.9% of the nodes (7893 nodes) and 98.6% of the edges (16,766 edges) from the reduced network.

Nodes (contributors) in the network spread across the continents, with higher concentrations in North and South Americas, Europe, South Africa, Southeast Asia, and Oceania (see Section 5.1 for details). The node degree follows a right-tailed distribution with values ranging from 1 to 326 (mean = 4; standard deviation = 11). Among the three edge attributes, a significant positive correlation exists between land cover composition similarity and species taxon composition similarity (Spearman's $r = 0.925$; p-value <0.001). However, no significant correlation was found between geographic distance and either land cover or species taxon composition similarity (see Section 5.3 for details).

4. The geovisual analytics tool

4.1. Overview

A custom geovisual analytics tool was developed to address the research questions laid out in Introduction regarding patterns and drivers of social interactions in iNaturalist (and other VGI communities). The tool provides means for exploring geographic/non-geographic patterns of inter-contributor species identification social interactions in iNaturalist (and other spatial social networks) and examining associations between social interaction and geographic contextual factors, represented by node/edge attributes. However, it is important to note that the tool primarily serves to facilitate exploratory data analysis rather than to support end-to-end full-cycle data analysis, although it can provide valuable insights that inform subsequent formal investigations, such as rigorous statistical analyses (Andrienko and Andrienko 2006).

The geovisual analytics tool was implemented as a single web page provisioned through a standard web server. Specifically, the tool offers interactive visualization and analytics capabilities to support exploring spatial social networks in both network domain and geographic domain where geographic context can be integrated (Figure 2). On the one hand, the tool extends beyond providing 2D/3D standard (non-geographic) social network visualizations by situating the network within geographic space, superimposing it on 2D background base maps or a 3D globe based on nodes' geographic locations. This facilitates interaction between the social network and its geographic environment because spatial proximity among nodes is eminently visualized and characteristics of the geographic environments (e.g. topography and land cover basemaps) where the nodes are embedded are readily available for visual examination. Network views in 3D and on the globe are provided as supplementary visualizations to the 2D views, as they can be more helpful in some cases (e.g. reducing the distortion of distance between nodes on a global scale or facilitating the visualization of node clustering in 3D on large networks through rotational exploration). On the other hand, beyond supporting classic social network analysis methods (PageRank analysis, connected component analysis, community detection, etc.) (Bedi and Sharma 2016; Oliveira and Gama 2012), it provides analytics functions to visualize and summarize the distributional characteristics of node/edge attributes, including those derived to represent interaction intensity (edge weight, i.e. number of species identification interactions) and geographic contextual factors (e.g. geographic distance and land cover composition). Patterns of the interactions could be extracted based on the distributions. The tool also supports examining correlations among node and edge attributes to reveal potential associations between interaction

	Visualization	Analytics

Network domain	• Social network visualization in non-geographic space (2D/3D)
	  • Attribute-based node rendering (shape, size, color, etc.)
	  • Weight based edge rendering (width, flow animation speed, etc.)
		• Centrality (node degrees, PageRank, etc.)
		• Connected component analysis
		• Community detection
		• Distribution of network metrics (degree, PageRank score, etc.)
		• Correlation among the metrics

Geographic domain	• Base maps (2D)
	  • Topography
	  • OpenStreetMap
	  • Satellite imagery
	  • Land cover
	• Social network visualization in geographic space (2D /3D)
		• Distribution of node/edge attributes
		  • Geographic distance
		  • Land cover composition (similarity)
		  • Species taxon composition (similarity)
		• Correlation between edge weight (i.e., interaction intensity) and the attributes

Figure 2. Geographic context-aware visualization and analytics functionalities provided by the geovisual analytics tool.

intensity and the geographic contextual factors. The tool allows for various filtering options (e.g. by distance between nodes; see section 4.2.1 for details). Any filtering will result in a 'focus' sub-network being extracted from the network for visualization and analysis. All visualization and analytics panels are linked and updated together on the 'focus' network.

The visual interfaces provided by the geovisual analytics tool are presented in this section. A small sub-network extracted from the iNaturalist network (through community detection) was used here for demonstration purposes. This network consists of 119 nodes and 192 edges with two connected components. The tool with this demo network is available at https://guiming.github.io/SocialNetworkVGI/ for interested readers to play with.

4.2. Network visualizations

The geovisual analytics tool (Figure 3) provides different network visualizations to allow the user to interactively explore the network: the network on 2D map (primary view; Figure 3), the network on 3D globe, and traditional non-geographic 2D or 3D network visualizations (secondary views; Figure 4). The traditional network

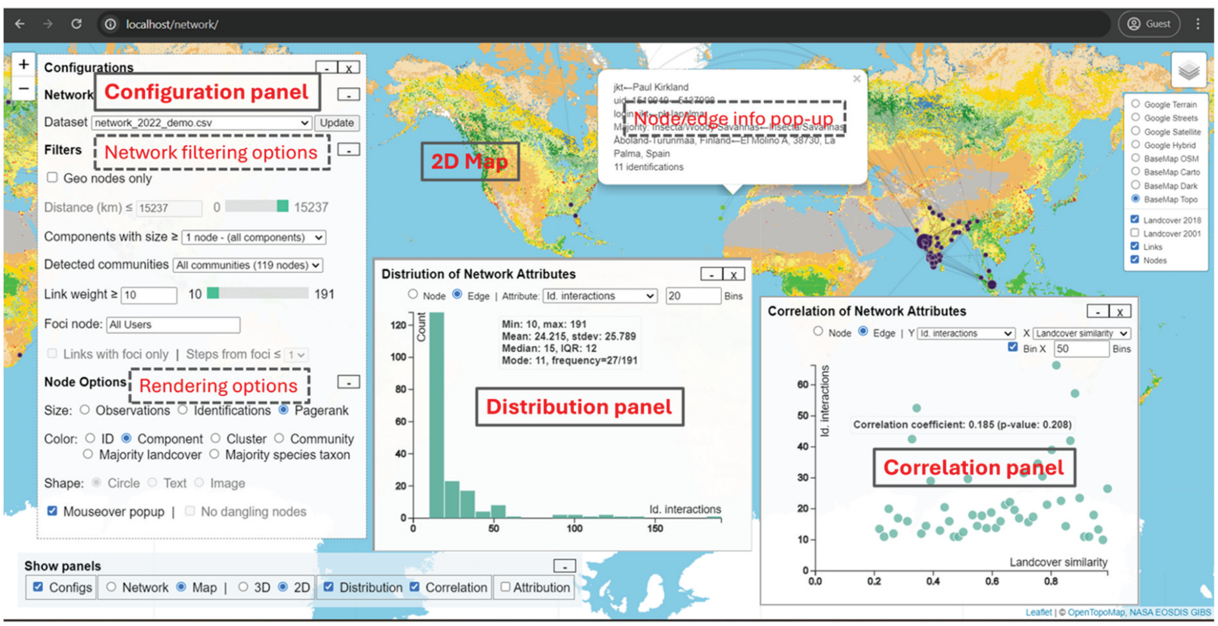

Figure 3. Overview of the geovisual analytics tool featuring an interactive spatial network visualization situated on a 2D land cover base map. A pop-up window shows node/edge information when the mouse hovers over a node/edge. The configuration panel allows users to set up network filtering options (e.g. distance between nodes) to extract a sub-network (as the 'focus' network) for visualization and analysis, as well as change visualization rendering options. The distribution panel provides bar charts showing the frequency distribution of numeric node/edge attributes specified by the user (or pie charts showing the composition of categorical attributes). The correlation panel offers scatter plots of two user-selected node/edge attributes accompanied by correlation coefficient and p-value. All panels are draggable and float on the map. They can be turned off through the show panels control, allowing a user to focus on the network and map.

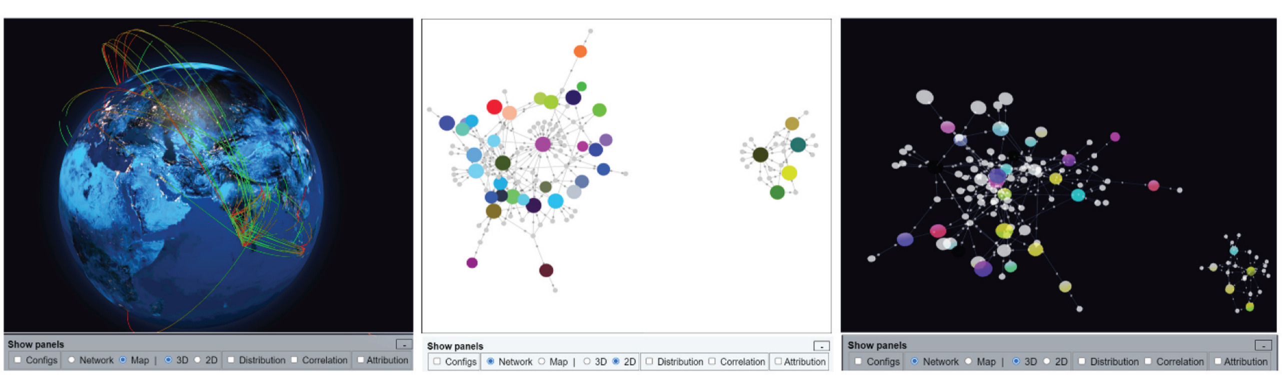

Figure 4. Additional interactive network visualizations provided by the geovisual analytics tool: a network on 3D globe (left), a 2D network (center), and a 3D network (right) in non-geographic space.

visualizations are good for examining relative node position in the network space (e.g. network distance between nodes), while the geographic visualizations of the network readily reveal geographic proximity among nodes and spatial clustering of the nodes. For instance, the 2D/3D non-geographic network visualizations immediately show two isolated components in the demo network (Figure 4 centre and right) and the geographic network visualizations (Figures 3 and 4 left) clearly show most nodes are clustered in India. The user can also choose a topographic map, street map, satellite imagery, and land cover map as the geographic background to situate the spatial network to visually explore characteristics of the places the nodes are tied to.

The tool has a configuration panel with various filtering and rendering options, enabling users to filter nodes/edges in the network and visualize nodes (Figure 3). The user may apply filters to exclude nodes without geographic location, edges whose geographic distance exceeds a certain threshold, or edges with weights below a threshold. If there are multiple connected components or communities detected in the network, the user can select which component or community to visualize (and analyse). The user can also limit the network to nodes that are directly connected to a foci node of interest or indirectly connected within a specified number of network steps. In these cases, the user is able to focus on a subset of the network (e.g. a component or community or a sub-network involving a specific node). Regarding rendering, nodes can be rendered as simple circles, texts (names), or images (profile pictures). Node colour can be assigned based on categorical node attributes, such as node id, community id, majority land cover, while node size can be scaled based on numeric node attributes, such as number of species observations, number of species identifications, and PageRank score.

4.3. Analytics functions

The tool provides analytics functions to support exploring patterns in the social network. In addition to basic network analysis, such as computing node degrees and PageRank scores and detecting components and communities in the network, the tool offers custom analytics functions to examine the distribution characteristics of node/edge attributes and investigate potential correlations between the attributes to facilitate understanding geographic context's influence on social interactions. The distribution panel (Figure 3) visualizes the frequency distribution of node attributes (number of species observations, number of species identifications, degree, PageRank score, land cover composition, and species taxon composition) and edge attributes (edge weight, geographic distance, land cover composition similarity, and species taxon composition similarity). Numeric attributes are displayed as histogram bar charts, while categorical attributes are represented as composition pie charts. The Correlation panel (Figure 3) allows users to explore correlations between attributes using scatter plots, accompanied by correlation coefficient and p-value. Examining the distribution of node/edge attributes could uncover patterns of social interactions in the network, while investigating correlations between the attributes helps shed light on the social and geographic factors influencing these patterns. For example, the bar chart in Figure 3 shows that edge weight (number of species identified between an observer-identifier pair) follows a long-tail distribution, meaning that most of the observer-identifier pairs (edges) have small numbers of identification interactions, while only a small portion of the pairs has large quantities of identification interactions. The scatter plot in Figure 3 shows a positive correlation (although statistically insignificant) between edge weight and land cover composition similarity), reflecting a general tendency that identifiers tend to identify more species observations from observers whose land cover composition of observation locations (i.e. observer land cover composition) is more similar to the land cover composition of the observations identified by the identifier (i.e. identifier land cover composition). By interacting with the network visualizations (including altering the filtering and rendering options) and examining the distribution graphs and correlation scatter plots, the user can explore and discover patterns and associations in the social network. These insights can help users formulate hypotheses about the social and geographic contextual drivers underlying these patterns.

5. Geovisual analysis of the iNaturalist network

The custom tool was utilized to explore the patterns of social interactions (i.e. observer-identifier species identification activities) within the iNaturalist network and the detected communities therein (Sections 5.1 and 5.2, respectively). Additionally, it was used to understand the social and geographic contextual factors that may have shaped the interaction patterns within the iNaturalist community (Section 5.3).

5.1. The iNaturalist network

Contributors in the iNaturalist network spread across the globe, with more contributors in North/South America, Europe, South Africa, South/Southeast/East Asia, and Oceania (Figure 5). However, inter-contributor species identification interactions in iNaturalist are not limited among

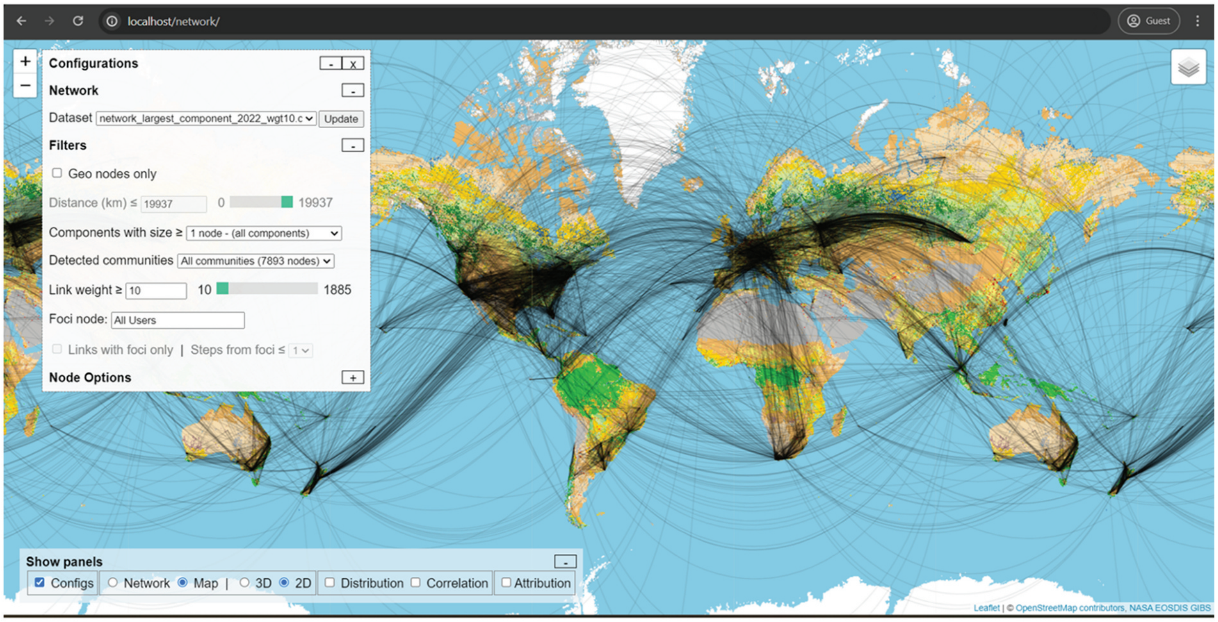

Figure 5. A geovisualization of the iNaturalist spatial social network (7893 nodes and 16,766 edges).

contributors in the same or nearby regions. This is evident from the edges linking contributors located in distant parts of the world. Patterns in the nodes and edges of the network are examined in the following sections.

5.1.1. The node attributes

Node attributes of the network, namely number of species observations, number of species identifications, node degree (number of connected nodes), and PageRank score (a centrality measure), all follow a long-tail distribution that skews to the right. Among the contributors, the majority of them (5736 out of 7893; 72.7%) contributed only one species identification, and many submitted relatively few species observations (median = 425 observations) (Figure 6). In contrast, a ew number of contributors contributed many identifications (maximum = 20,147 identifications) and/or observations (maximum = 27,122 observations). The network is sparsely connected, with over 50% of the contributors (3954 out of 7893) connected to only one other node in the network, and these nodes have very low PageRank scores. In contrast, a few contributors are connected to many nodes (maximum node degree = 326) and are associated with high PageRank scores. This network structure may result from a preferential attachment process (McCulloh, Armstrong, and Johnson 2013), wherein a small number of contributors play central roles in contributing species identifications and/or observations. These individuals, who have a higher number of connections, are more 'popular' or well-known, making them more likely to gain new connections.

5.1.2. The edge attributes

Examining edges in the network (Figure 7), edge weight (i.e. number of species identifications between an observer-identifier pair) also follows a long-tail distribution with a rightward skew. This indicates that most observer–identifier interactions are associated with a small number of species identifications (half of the interactions have no more than 15 identifications), while only a few interactions involve large quantities of identifications (maximum = 1885 identifications). Geographic distance also follows a right-skewed long-tail distribution, indicating that species identification interactions tend to occur between contributors who are geographically closer. In contrast, land cover composition similarity and species taxon composition similarity both follow left-skewed long-tail distributions. This implies that interactions tend to occur between observer-identifier pairs whose land cover or species taxon compositions (see Section 3) are more.

5.2. Communities within the network

In a social network, actors with similar backgrounds, preferences, and/or interests often associate to form virtual clusters or communities (McCulloh, Armstrong, and Johnson 2013). That is, nodes within the same community have more interactions with each other than with nodes outside the community (Javed et al. 2018). Using the geovisual analytics tool, which implemented the Louvain community detection algorithm (Blondel

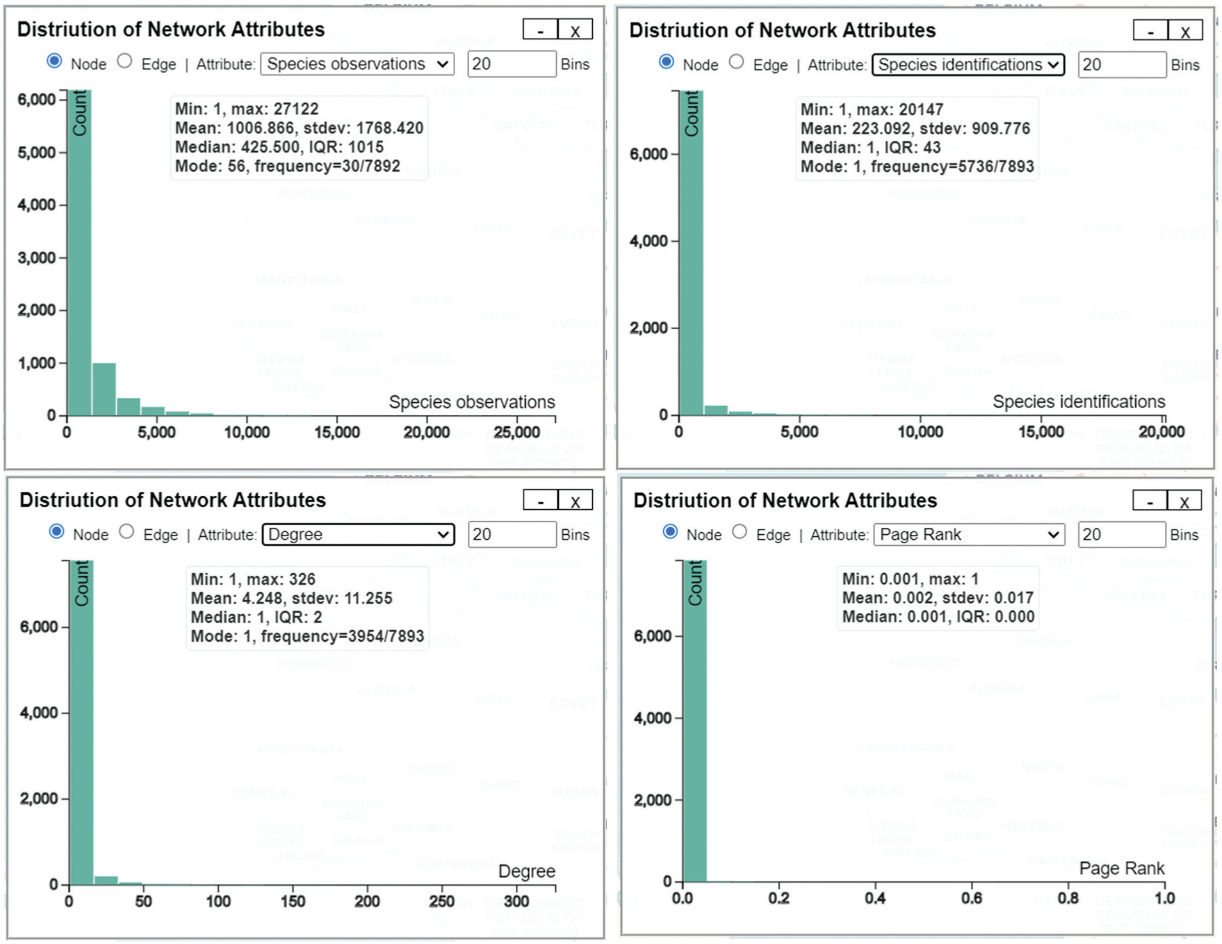

Figure 6. Distribution of node attributes: the number of species observations, the number of species identifications, node degree, and PageRank score.

et al. 2008), a total of 34 communities were detected within the iNaturalist network. A map of the communities (Figure 8) reveals that nodes within a community tend to cluster in geographic space, indicating that iNaturalist contributors appear to interact more with other contributors who are geographically close (i.e. identifiers are more likely to identify species observations submitted by 'nearby' observers).

Among the communities detected within the iNaturalist network, 11 communities consisted of over 200 nodes (Table 1). These communities were further profiled using a geovisual analytics tool to discover patterns in the communities. Most contributors in the same community appear to be geographically clustered within a region or country. For example, while the largest community (Community 0) spreads across North America, Community 1 is centred in the Eastern US, and Community 8 is centred in Canada (Figure 9). Compared to the iNaturalist network, 10 out of 11 communities have a greater (or equal) median edge weight, meaning that observer-identifier pairs in these respective communities tend to interact with higher intensities (i.e. identifying more species observations on average). Eight communities have a lower median geographic distance, indicating that contributors in the communities overall are geographically closer than those in the iNaturalist network. Seven communities have a higher median species taxon composition similarity, and nine communities have a higher median land cover composition similarity, implying that, compared to the iNaturalist network, contributors in the communities generally share a greater interest in certain species taxon kingdoms and/or a stronger preference on observation environment, as represented by land cover type.

Additionally, communities appear to have distinct interests in specific species taxon kingdoms and/or preferences on certain observation environments. For example, compared to the baseline (contributors in the full iNaturalist network), contributors in Community 6 (New Zealand) are more interested in identifying species observations in the Actinopterygii kingdom (ray-finned fishes) (Figure 10; Table 2) and observations from water

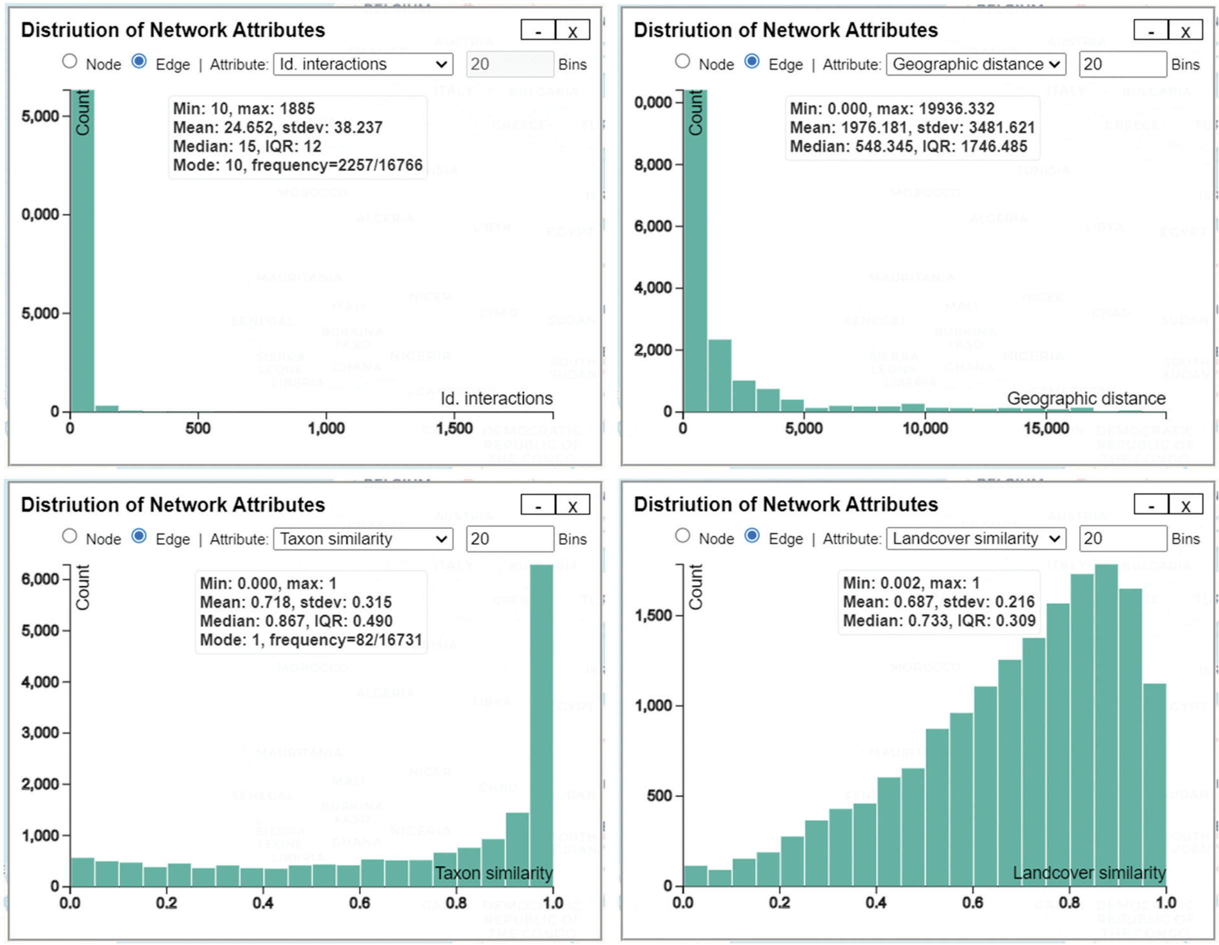

Figure 7. Distribution of edge attributes: edge weight, geographic distance, land cover composition similarity, and species taxon composition similarity.

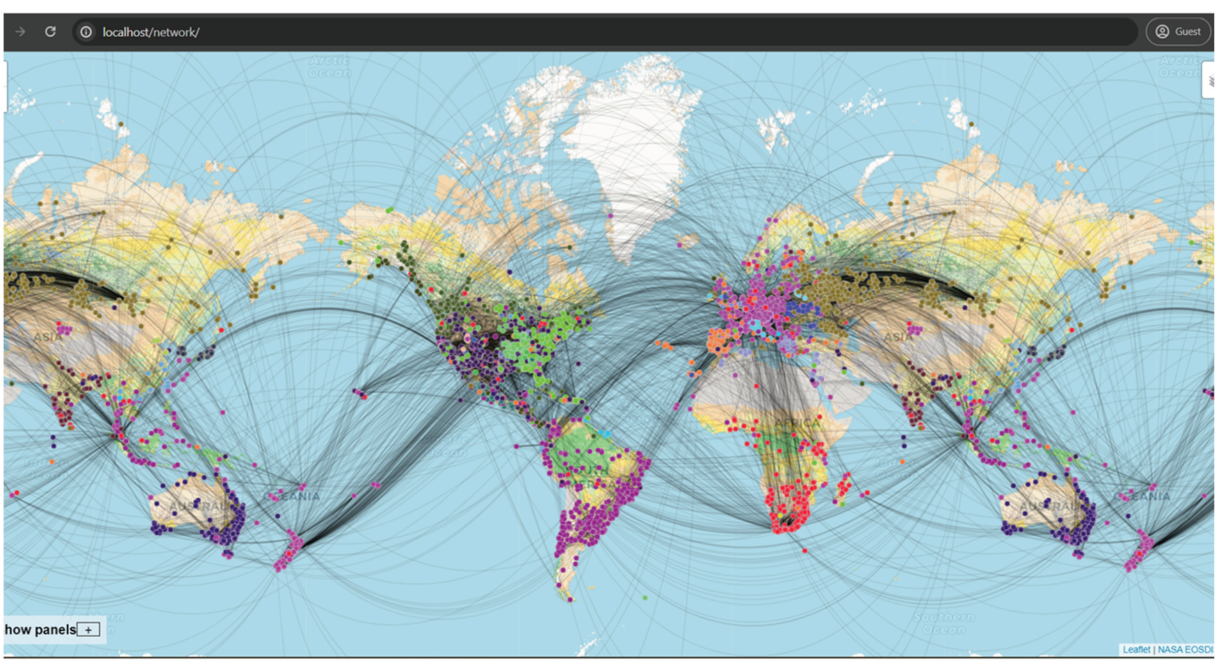

Figure 8. The iNaturalist social network with node color representing the communities detected in the network.

Table 1. Summary statistics for the iNaturalist network and the communities detected within the network (C0 through C10).

	iNaturalist	C0	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Number of nodes	7893	1357	818	763	692	680	631	501	457	393	322	298
Median edge weight	15	14	**16**	**16**	**16**	**18**	**16**	**16.5**	**16**	15	**19**	15
Median geographic distance	548	**475**	**283**	**445**	714	805	**536**	646	**412**	**254**	**98**	**311**
Median taxon composition similarity	0.87	**0.89**	0.85	**0.89**	0.78	**0.98**	**0.90**	0.79	0.80	**0.90**	**0.95**	**0.92**
Median land cover composition similarity	0.73	**0.77**	**0.74**	0.71	**0.77**	**0.74**	0.73	**0.74**	**0.74**	**0.77**	**0.84**	**0.79**

For each community, values that are higher than the iNaturalist network median edge weight, shorter geographic distance, higher species taxon similarity composition similarity, and higher land cover composition similarity are highlighted in bold.

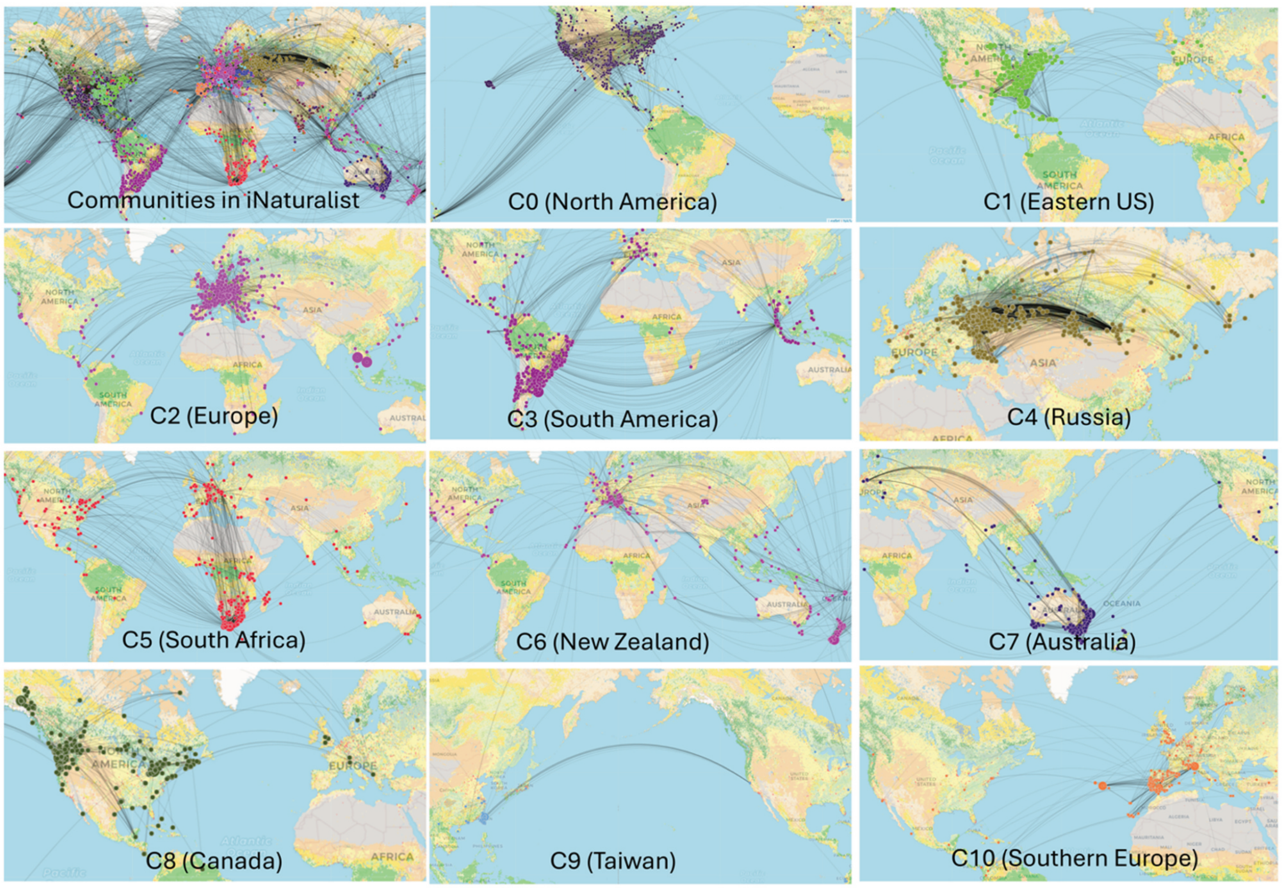

Figure 9. The eleven largest communities (C0 through C10) detected in the iNaturalist social network. Each community is labeled loosely based on the geographic region where most community members are located.

bodies (Figure 11; Table 3). Communities 0 (North America) and 5 (South Africa) both are more interested in identifying bird observations. Also, communities in North America focus on observations made in urban and built-up environments, while those in South Africa focus on observations made on grasslands. Communities 1 (Eastern US), 2 (Europe), 3 (South America), and 7 (Australia) are all more interested in insect observations and observations from woody savannas, croplands, urban and built-up lands, and evergreen broadleaf forests, respectively. Communities 4 (Russia), 8 (Canada), 9 (Taiwan), and 10 (Southern Europe) are all more interested in plant observations and observations from mixed forests, evergreen needleleaf forests, evergreen broadleaf forests, and grassland, respectively

5.3. Drivers of interactions

To explore potential drivers of species identification interactions in iNaturalist, the associations between edge weight that indicates interaction intensity (i.e. number of inter-contributor species identifications) and

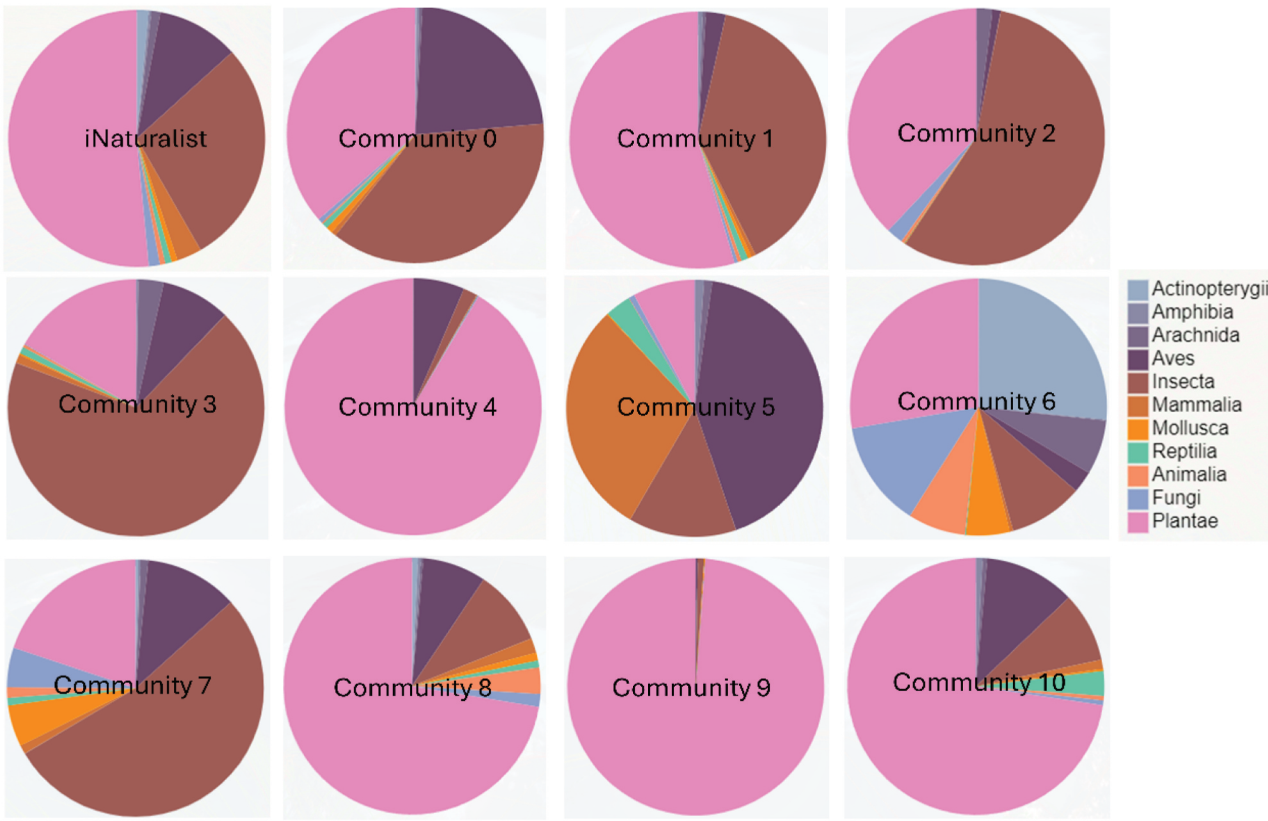

Figure 10. Species taxon composition of identified observations aggregated for the iNaturalist network and for communities detected within the network.

Table 2. Aggregated identifier species taxon composition in the iNaturalist network (baseline) and deviations from the baseline composition for communities detected in the network (C0 through C10).

	Baseline (%)	Deviation from baseline (%)										
	iNaturalist	C0	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Actinopterygii	1.5	−1.4	−1.4	−1.5	−1.5	−1.5	−1.5	**25.0**	−1.1	−0.8	−1.5	−1.5
Amphibia	0.3	0.1	0.2	−0.3	−0.1	−0.3	0.8	−0.2	−0.1	0.1	−0.3	0.5
Arachnida	1.1	−0.7	−0.7	0.9	2.0	−1.1	−0.1	5.7	−0.2	−0.8	−1.1	−0.5
Aves	10.3	**12.4**	−7.8	−9.3	−1.6	−3.9	**32.3**	−7.6	1.4	−2.3	−10.1	1.1
Insecta	28.4	8.6	**10.7**	**27.9**	**40.2**	−26.7	−14.8	−19.1	**24.8**	−18.9	−27.6	−19.6
Mammalia	3.2	−2.5	−2.6	−3.1	−2.1	−3.0	26.5	−2.7	−2.0	−1.3	−3.1	−2.0
Mollusca	0.7	0.1	−0.3	−0.7	−0.5	−0.7	−0.6	4.8	4.5	0.2	−0.6	−0.5
Reptilia	0.8	−0.1	0.0	−0.8	0.1	−0.8	2.4	−0.7	0.1	0.0	−0.8	2.3
Animalia	0.7	−0.5	−0.2	−0.2	−0.3	−0.7	−0.6	6.6	0.7	2.7	−0.7	−0.1
Fungi	1.4	−0.9	−1.0	0.7	−1.3	−1.3	−0.7	12.0	3.6	0.2	−1.4	−0.8
Plantae	51.5	−15.0	3.1	−13.5	−34.8	**40.0**	−43.7	−23.8	−31.6	**20.9**	**47.3**	**21.1**

For each community, the species taxon kingdom with the highest positive deviation is highlighted in bold.

three edge attributes representing geographic context were examined using the geovisual analytics tool. In the iNaturalist network, edge weight is negatively correlated with geographic distance (correlation coefficient = −0.406; $p < 0.1$), positively with species taxon composition similarity (correlation coefficient = 0.788; $p < 0.01$), and positively with land cover composition similarity (correlation coefficient = 0.871; $p < 0.01$) (Figure 12). The correlations suggest that overall species identification interactions involving more observations tend to occur between observers and identifiers who are geographically closer, have more similar taxonomic interests in species, and/or share more similar preferences on observation environments (land cover composition). Therefore, the three geographic contextual factors, geographic distance, land cover composition similarity, and species taxon composition similarity – appear to be drivers of species identification interactions in the iNaturalist network.

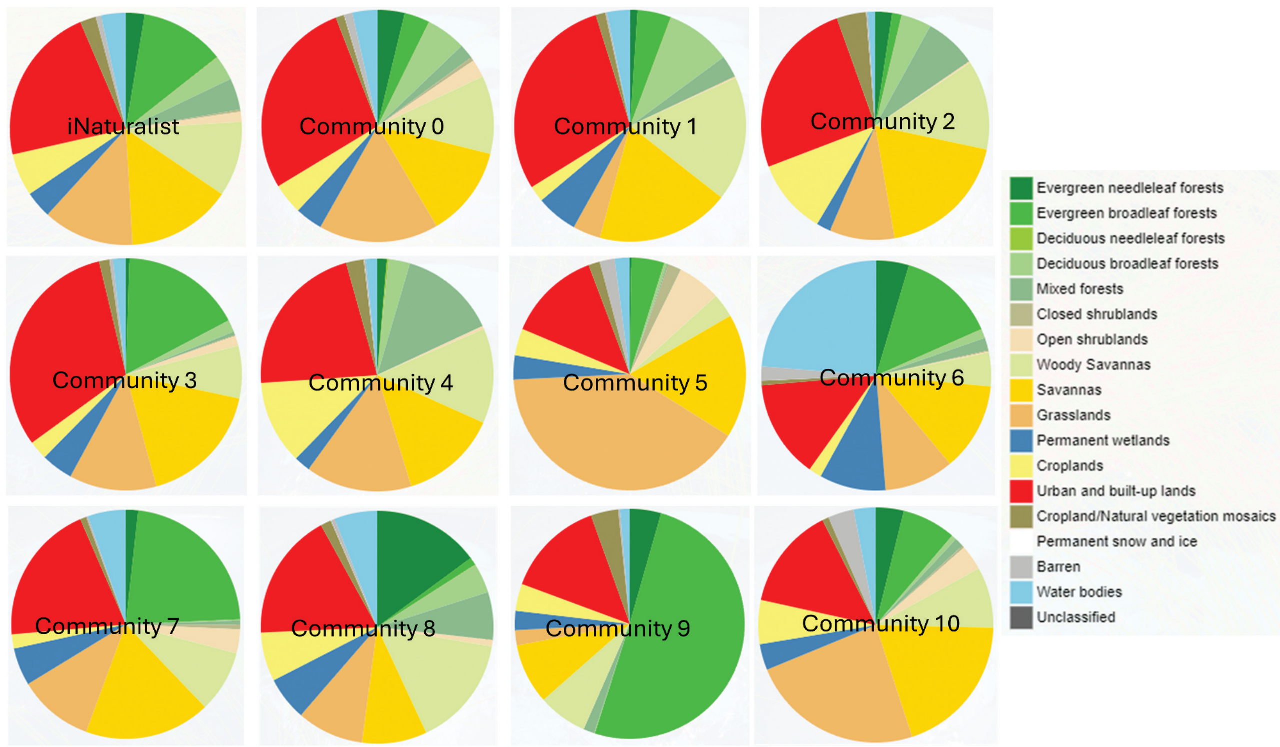

Figure 11. Land cover composition of identified observations aggregated for the iNaturalist network and for communities detected within the network.

Table 3. Aggregated identifier land cover composition in the iNaturalist network (baseline) and deviations from the baseline composition for communities detected in the network (C0 through C10).

	Baseline (%)	Deviation from baseline (%)										
	iNaturalist	C0	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Evergreen needleleaf forests	2.6	1.2	−1.6	−0.2	−2.1	−1.3	−2.4	2.0	−0.9	**12.3**	1.8	1.2
Evergreen broadleaf forests	11.7	−8.2	−7.1	−10.4	5.1	−11.7	−7.2	2.0	**10.8**	−10.7	**38.8**	−4.4
Deciduous needleleaf forests	0.1	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0
Deciduous broadleaf forests	3.5	2.1	5.7	0.8	−1.8	−0.5	−3.1	−2.2	−3.3	0.6	−3.4	−2.8
Mixed forests	4.4	−2.3	−1.3	3.0	−3.9	**9.1**	−4.3	−2.6	−3.8	2.2	−2.8	−3.1
Closed shrublands	0.3	0.3	−0.2	−0.2	−0.2	−0.3	1.6	−0.1	0.4	−0.2	−0.3	0.0
Open shrublands	1.4	1.1	−1.2	−1.2	0.0	−1.0	4.6	−1.2	2.0	−0.5	−1.4	2.1
Woody Savannas	10.5	0.4	**6.9**	2.2	−3.1	2.7	−7.1	−5.7	−1.7	5.0	−3.6	−2.1
Savannas	14.6	−1.9	4.0	4.3	2.8	−1.1	2.6	−2.2	3.2	−5.6	−6.1	5.0
Grasslands	12.7	4.0	−8.8	−3.5	−0.5	2.1	**27.7**	−2.9	−2.1	−3.4	−10.6	**11.1**
Permanent wetlands	3.8	0.1	2.1	−1.7	0.7	−1.5	−0.4	5.6	1.6	2.3	−1.1	−0.2
Croplands	5.9	−1.6	−3.7	**4.8**	−3.3	5.7	−2.2	−4.1	−3.9	0.9	−2.1	0.1
Urban and built-up lands	22.1	**5.7**	**6.9**	3.2	**9.2**	−0.3	−9.2	−8.3	−2.2	−4.4	−8.2	−8.0
Cropland/Natural vegetation mosaics	2.2	−1.1	−0.9	2.0	−0.8	0.2	−0.6	−1.6	−1.4	−0.7	1.6	−1.3
Permanent snow and ice	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.6
Barren	0.7	0.4	−0.5	−0.5	−0.3	−0.5	1.3	1.2	−0.5	−0.2	−0.5	2.3
Water bodies	3.4	0.1	−0.2	−2.4	−1.6	−1.9	−1.4	**20.3**	1.9	2.5	−2.1	−3.4

For each community, the land cover type with the highest positive deviation is highlighted in bold.

A caveat is that a statistically significant positive correlation (correlation coefficient = 0.925; $p < 0.01$) was observed between land cover composition and species taxon composition similarities. This is somewhat expected given that there are strong associations between species and their habitats (e.g. land cover) (Figure 13). However, the similarities are not correlated with geographic distance. The implication is that, although contributors' common preferences for certain observation environments (as measured by land cover composition similarity) and shared interests on certain species taxon (as measured by taxon composition similarity) are inter-dependent and may not independently drive inter-contributor species identification interactions in iNaturalist, geographic distance remains an independent driver of the interactions.

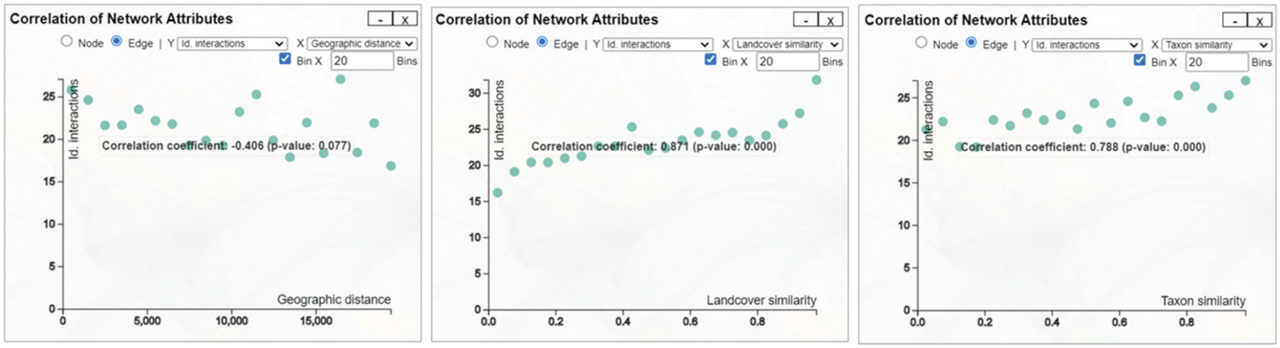

Figure 12. Correlation between edge weight (the number of inter-contributor species identifications) and geographic distance (left), land cover composition similarity (center), and species taxon composition similarity (right).

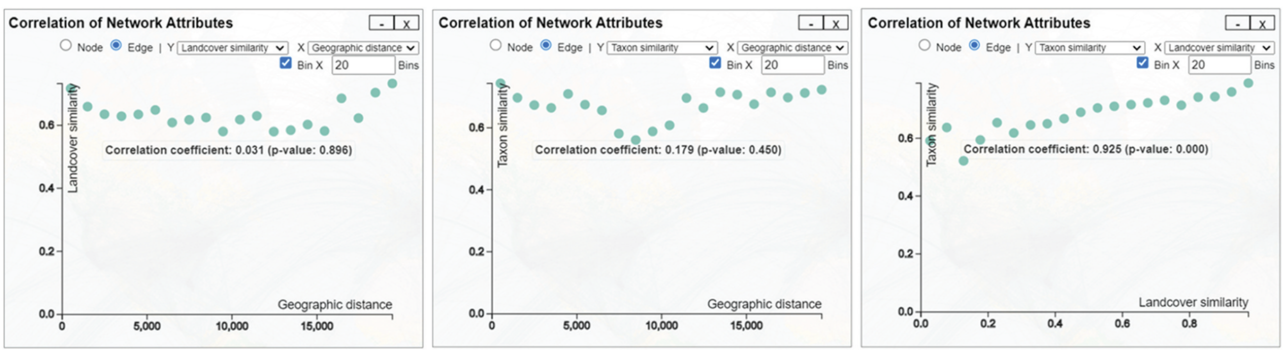

Figure 13. Co-linearity among the three edge attributes. Correlations between geographic distance and land cover composition similarity (left), geographic distance and species taxon composition similarity (center), and land cover composition similarity and species taxon composition similarity (right).

Table 4. Correlation coefficients between edge weight and geographic distance, species taxon composition similarity, and land cover composition similarity in the communities detected within the network (C0 through C10).

	C0	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Geographic distance	−0.18	0.15	**−0.51***	−0.07	0.04	−0.35	−0.09	−0.37	−0.33	−0.32	−0.27
Taxon composition similarity	**0.63*****	0.02	0.07	0.04	**0.63*****	0.27	**0.50****	−0.10	0.11	**0.46****	0.36
Land cover composition similarity	**0.50****	**0.61*****	**0.52****	**0.62*****	0.29	−0.08	0.31	0.12	0.23	**0.67*****	−0.07

Statistically significant correlation coefficients at the 0.1 level are highlighted in bold. *** p value < 0.01, ** p value < 0.05, * p value < 0.1.

Similar correlations were observed within the communities detected in the iNaturalist network. In six of the 11 communities, at least one of the three coefficients is statistically significant (Table 4). That is, species identification interaction intensity within these communities is significantly correlated with at least one of the three geographic contextual factors (i.e. geographic distance, land cover composition similarity and species taxon composition similarity). Moreover, as noted earlier (Table 1; section 5.2), observer-identifier pairs (i.e. edges) in the respective communities have a shorter median geographic distance, higher median land cover composition similarity and higher species taxon composition similarity compared to the iNaturalist network. These findings indicate that the three geographic contextual factors may have influenced both the formation of the communities and the intensity of species identification interactions within them. This further supports the observation that these geographic contextual factors are likely among the key drivers of species identification interactions within the iNaturalist community.

The findings on potential drivers of social interactions in the iNaturalist community align well with existing theories concerning the key forces driving the formation of social interactions in a social network. These theories identify proximity, homophily, reciprocity, prestige, transitivity, social conformity, and balance as the key social forces driving interactions in social networks (McCulloh, Armstrong, and Johnson 2013) (see Introduction). Among these forces, proximity refers to the physical distance between social actors, with closer actors being more likely to interact than those farther apart. Homophily describes the tendency of actors with common interests, preferences, beliefs, etc. to establish social interactions. In the context of iNaturalist, geographic distance as a driver of social interaction reflects the effect of physical proximity, while species taxon composition similarity and land cover composition similarity reflect homophily effects.

6. Discussion and conclusion

This study explores social interaction patterns within the iNaturalist social network using a custom geovisual analytics tool. The social network was constructed based on the inter-contributor species identification activities within the iNaturalist community. Three edge attributes (i.e. geographic distance, land cover composition similarity, and species taxon composition similarity) were derived to represent geographic contextual factors that may impact social interactions in the iNaturalist network. Geovisual explorations of the network revealed that a small number of contributors play central roles in contributing species identifications and/or observations. Moreover, species identification interactions in iNaturalist tend to occur more frequently and/or at higher intensity between contributors who are geographically closer, who have more common interests in certain species taxon (as represented by species taxon composition similarity), and/or who have more similar preferences on certain types of observation environments (as represented by land cover composition similarity). These factors also seem to influence the formation of the communities within the network as well as impact the interaction intensity within the communities. These findings from geovisual explorations of the iNaturalist network support the observation that the three geographic contextual factors are among the drivers of species identification interactions within the iNaturalist community. This observation aligns with the theories concerning the key social forces driving the formation of social interactions in a social network (McCulloh, Armstrong, and Johnson 2013), wherein geographic distance reflects physical proximity and land cover composition similarity and species taxon composition similarity reflect homophily effects.

The custom geovisual analytics tool is capable of visualizing and analysing spatial social networks in geographic context for exploring patterns of social interactions within VGI communities and thus effectively addresses the disconnect between geographic context-aware visual network exploration and analysis. The tool integrates geographic context into social network visualization and analysis at two levels. First, the tool provides maps depicting the geographic environments, wherein the social network is embedded, as background base maps. The network is then superimposed on top of these base maps to offer geographic context-explicit network visualizations. Second, it incorporates node and edge attributes reflecting geographic contextual factors (geographic location, land cover, etc.) into network analysis and offers analytics functions for exploring associations between social interaction intensity and the contextual factors. The tool effectively facilitates exploring patterns and drivers of social interactions and offers a new perspective to examine the social and geographic dynamics of social interactions within VGI communities. This tool could also potentially facilitate developing hypotheses for understanding potential drivers of spatial social interactions through the lens of similarities in geographical, social, and semantic spaces (Luo et al. 2019). Although mainly designed as an exploratory data analysis tool, the insights gained through geovisual exploration can inform the formulation of hypotheses regarding the patters and drivers of social interactions in a social network, which can be comprehensively tested in subsequent studies (Zhang, Gong, and Zhu 2024a). This study, along with the spatio-social-semantic analysis framework (Luo et al. 2019) for analysing spatial social networks, offers a theoretical foundation affirming that social networks and their social interactions need to be examined across different spaces.

This study has certain limitations. First, only land cover was adopted to depict the physical environment wherein iNaturalist contributors conduct observations, although geographic context is multifaceted. Other aspects of the environment (e.g. socioeconomic characteristics) may also influence a contributor's interaction with others. When characterizing geographic context and incorporating it into social network analysis, decisions regarding what contextual factors need to be considered and how to represent them in the social network (e.g. as node attributes and/or edge attributes) should be made based on domain-specific knowledge of potential influences on a social actor's behaviours, as well as on data availability. Second, iNaturalist is not necessarily representative of all the varieties of VGI communities although it reflects the characteristics of VGI projects where data contributor's interactions play an important role in VGI data creation and are influenced by geographic contexts. Third, the tool currently provides only basic statistical analyses (e.g. correlation analysis) and does not yet include spatial statistical analyses (e.g. testing spatial clustering of nodes). Fourth, edges in the original iNaturalist network with weights below the threshold (i.e. 10) were removed to reduce computational load.

Omitting these edges, which might still hold nontrivial information about occasional or less frequent interactions, may introduce bias and uncertainty in the analyses. Finally, from a technical standpoint, the geovisual analytics tool has limitations in handling very large networks consisting of many nodes and edges, as its rendering and computing capabilities are limited by the underlying JavaScript libraries that run on web browsers.

Acknowledgements

The authors appreciate participants of the iNaturalist biodiversity citizen science project for their generous efforts in contributing species observations that make this study possible. The APC was sponsored by the University of Denver's Open Access Publication Equity Fund.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the University of Denver through the Professional Research Opportunities for Faculty (PROF) Grant.

ORCID

Mingguang Wu http://orcid.org/0000-0001-6086-4503

References

Andrienko, N., and G. Andrienko. 2006. *Exploratory Analysis of Spatial and Temporal Data: A Systematic Approach*. Springer Berlin, Heidelberg: Springer Science & Business Media.

Andris, C. 2016. "Integrating Social Network Data into GISystems." *International Journal of Geographical Information Science* 30 (10): 2009–2031. https://doi.org/10.1080/13658816.2016.1153103.

Andris, C., D. DellaPosta, B. N. Freelin, X. Zhu, B. Hinger, and H. Chen. 2021. "To Racketeer Among Neighbors: Spatial Features of Criminal Collaboration in the American Mafia." *International Journal of Geographical Information Science* 35 (12): 2463–2488. https://doi.org/10.1080/13658816.2021.1884869.

Aristeidou, M., C. Herodotou, H. L. Ballard, A. N. Young, A. E. Miller, L. Higgins, R. F. Johnson, and C. A. Lepczyk. 2021. "Exploring the Participation of Young Citizen Scientists in Scientific Research: The Case of iNaturalist." *PLOS ONE* 16 (1): 1–14. https://doi.org/10.1371/journal.pone.0245682.

Bedi, P., and C. Sharma. 2016. "Community Detection in Social Networks." *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 6 (3): 115–135. https://doi.org/10.1002/widm.1178.

Blondel, V., J. Guillaume, R. Lambiotte, and E. Lefebvre. 2008. "Fast Unfolding of Community Hierarchies in Large Network, 2008." *Journal of Statistical Mechanics: Theory & Experiment* 2008 (10): 1008. https://doi.org/10.1088/1742-5468/2008/10/P10008.

Elwood, S., M. F. Goodchild, and D. Z. Sui. 2012. "Researching Volunteered Geographic Information: Spatial Data, Geographic Research, and New Social Practice." *Annals of the Association of American Geographers* 102 (3): 571–590. https://doi.org/10.1080/00045608.2011.595657.

Emch, M., E. D. Root, S. Giebultowicz, M. Ali, C. Perez-Heydrich, and M. Yunus. 2012. "Integration of Spatial and Social Network Analysis in Disease Transmission Studies." *Annals of the Association of American Geographers* 102 (5): 1004–1015. https://doi.org/10.1080/00045608.2012.671129.

Feng, Y., X. Huang, and M. Sester. 2022. "Extraction and Analysis of Natural Disaster-Related VGI from Social Media: Review, Opportunities and Challenges." *International Journal of Geographical Information Science* 36 (7): 1275–1316. https://doi.org/10.1080/13658816.2022.2048835.

Gao, S., H. Chen, W. Luo, Y. Hu, and X. Ye. 2018. "Spatio-Temporal-Network Visualization for Exploring Human Movements and Interactions in Physical and Virtual Spaces." In *Human Dynamics Research in Smart and Connected Communities*, edited by D.-S.-S.-L. Shaw, 67–80. Springer. https://doi.org/10.1007/978-3-319-73247-3_4.

Giordano, A., T. Cole, and M. Le Noc. 2022. "Spatial Social Networks for the Humanities: A Visualization and Analytical Model." *Transactions in GIS* 26 (4): 1683–1702. https://doi.org/10.1111/tgis.12938.

Gong, X., S. Peng, Y. Lu, S. Wang, X. Huang, and X. Ye. 2023. "Social Network Analysis of Nonprofits in Disaster Response: The Case of Twitter During the COVID-19 Pandemic in the United States." *Social Science Computer Review* 41 (6): 2029–2054. https://doi.org/10.1177/08944393221130674.

Gong, X., and X. Ye. 2021. "Governors Fighting Crisis: Responses to the COVID-19 Pandemic Across U.S. States on Twitter." *The Professional Geographer* 73 (4): 683–701. https://doi.org/10.1080/00330124.2021.1895850.

Goodchild, M. F. 2007. "Citizens as Sensors: The World of Volunteered Geography." *Geo Journal* 69 (4): 211–221. https://doi.org/10.1007/s10708-007-9111-y.

Haklay, M. 2021. "Geographic Citizen Science: An Overview." In *Geographic Citizen Science Design*, edited by A. Skarlatidou and M. Haklay, 15–37. London, UK.: UCLPRESS.

Haklay, M., and P. Weber. 2008. "OpenStreetmap: User-Generated Street Maps." *IEEE Pervasive Computing* 7 (4): 12–18. https://doi.org/10.1109/MPRV.2008.80.

Huang, X., S. Wang, D. Yang, T. Hu, M. Chen, M. Zhang, G. Zhang, et al. 2024. "Crowdsourcing Geospatial Data for Earth and Human Observations: A Review." *Journal of Remote Sensing* 4: 0105. https://doi.org/10.34133/remotesensing.0105.

iNaturalist. 2022. *iNaturalist Observations Export*. https://www.inaturalist.org/observations/export.

iNaturalist. 2024. *Frequently Asked Questions*. https://www.inaturalist.org/pages/help#quality.

Javed, M. A., M. S. Younis, S. Latif, J. Qadir, and A. Baig. 2018. "Community Detection in Networks: A Multidisciplinary Review." *Journal of Network and Computer Applications* 108 (September 2017): 87–111. https://doi.org/10.1016/j.jnca.2018.02.011.

Jin, S., A. Endert, and C. Andris. 2024. "SNoMaN: A Visual Analytic Tool for Spatial Social Network Mapping and Analysis." *Cartography and Geographic Information Science* 1–19. https://doi.org/10.1080/15230406.2024.2413600.

Jusup, M., P. Holme, K. Kanazawa, M. Takayasu, I. Romić, Z. Wang, S. Geček, et al. 2022. "Social Physics." *Physics Reports* 948:1–148. https://doi.org/10.1016/j.physrep.2021.10.005.

Kottwitz, M., G. Zhang, and J. Xu. 2023. "The Time- and Distance-Decay Effects of Hurricane Relevancy on Social Media: An Empirical Study of Three Hurricanes in the United States." *Annals of GIS* 29 (4): 469–484. https://doi.org/10.1080/19475683.2023.2236678.

Liu, Y. L., and T. Jiralerspong. 2023. "Network Analysis of the iNaturalist Citizen Science Community." https://doi.org/10.15468/dl.wt7szw.

Luo, W. 2016. "Visual Analytics of Geo-Social Interaction Patterns for Epidemic Control." *International Journal of Health Geographics* 15 (1): 1–16. https://doi.org/10.1186/s12942-016-0059-3.

Luo, W., and A. M. MacEachren. 2014. "Geo-Social Visual Analytics." *Journal of Spatial Information Science* 8 (8): 27–66. https://doi.org/10.5311/JOSIS.2014.8.139.

Luo, W., Y. Wang, X. Liu, and S. Gao. 2019. "Cities as Spatial and Social Networks: Towards a Spatio-Socio-Semantic Analysis Framework." In *Cities as Spatial and Social Networks. Human Dynamics in Smart Cities*, edited by X. Ye and X. Liu, 21–37. Cham,: Springer. 10.1007/978-3-319-95351-9_3.

Luo, W., P. Yin, Q. Di, F. Hardisty, A. M. MacEachren, and B. Jiang. 2014. "A Geovisual Analytic Approach to Understanding Geo-Social Relationships in the International Trade Network." *PLoS One* 9 (2): e88666. https://doi.org/10.1371/journal.pone.0088666.

McCulloh, I., H. Armstrong, and A. Johnson. 2013. *Social Network Analysis with Applications*. Hoboken, New Jersey: John Wiley & Sons.

Mesaglio, T., and C. T. Callaghan. 2021. "An Overview of the History, Current Contributions and Future Outlook of iNaturalist in Australia." *Wildlife Research* 48 (4): 289–303. https://doi.org/10.1071/WR20154.

Mooney, P., and P. Corcoran. 2014. "Analysis of Interaction and Co-Editing Patterns Amongst OpenStreetmap Contributors." *Transactions in GIS* 18 (5): 633–659. https://doi.org/10.1111/tgis.12051.

Oliveira, M., and J. Gama. 2012. "An Overview of Social Network Analysis." *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 2 (2): 99–115. https://doi.org/10.1002/widm.1048.

Quinn, S. D., and A. M. MacEachren. 2018. "A Geovisual Analytics Exploration of the OpenStreetmap Crowd." *Cartography and Geographic Information Science* 45 (2): 140–155. https://doi.org/10.1080/15230406.2016.1276479.

Radil, S. M., C. Flint, and G. E. Tita. 2010. "Spatializing Social Networks: Using Social Network Analysis to Investigate Geographies of Gang Rivalry, Territoriality, and Violence in Los Angeles." *Annals of the Association of American Geographers* 100 (2): 307–326. https://doi.org/10.1080/00045600903550428.

Robinson, A. 2017. "Geovisual Analytics." In *The Geographic Information Science & Technology Body of Knowledge*, edited by J. P. Wilson. https://doi.org/10.22224/gistbok/2017.3.

Sarkar, D., and J. T. Anderson. 2022. "Corporate Editors in OpenStreetmap: Investigating Co-Editing Patterns." *Transactions in GIS* 26 (4): 1879–1897. https://doi.org/10.1111/tgis.12910.

Sarkar, D., C. Andris, C. A. Chapman, and R. Sengupta. 2019. "Metrics for Characterizing Network Structure and Node Importance in Spatial Social Networks." *International Journal of Geographical Information Science* 33 (5): 1017–1039. https://doi.org/10.1080/13658816.2019.1567736.

Sbrocchi, C., G. Pecl, I. van Putten, and P. Roetman. 2022. "A Citizen Science Community of Practice: Relational Patterns Contributing to Shared Practice." *Citizen Science: Theory and Practice* 7 (1): 1–14. https://doi.org/10.5334/CSTP.358.

Schöttler, S., Y. Yang, H. Pfister, and B. Bach. 2021. "Visualizing and Interacting with Geospatial Networks: A Survey and Design Space." *Computer Graphics Forum* 40 (6): 5–33. https://doi.org/10.1111/cgf.14198.

Sulla-Menashe, D., and M. A. Friedl. 2018. "User Guide to Collection 6 MODIS Land Cover (MCD12Q1 and MCD12C1) Product."

Tupikina, L., F. Schlosser, V. Voskresenskii, K. Kloppenborg, F. Lopez, A. Mariz, A. Mogilevskaja, M. Haklay, and B. G. Tzovaras. 2021. "iNaturalist Citizen Science Community During City Nature Challenge: New Computational Approach for Analysis of User Activity. *ArXiv Preprint ArXiv:2112.02693*.

Ueda, K. 2022. "iNaturalist Research-Grade Observations. iNaturalist.Org. Occurrence Dataset." https://doi.org/10.15468/ab3s5x. accessed via GBIF.org.

Unger, S., M. Rollins, A. Tietz, and H. Dumais. 2021. "iNaturalist as an Engaging Tool for Identifying Organisms in Outdoor Activities." *Journal of Biological Education* 55 (5): 537–547. https://doi.org/10.1080/00219266.2020.1739114.

Wang, S., W. Li, and Z. Gu. 2023. "GeoGraphviz: Geographically Constrained 3D Force-Directed Graph for Knowledge Graph Visualization." *Transactions in GIS* 27 (4): 931–948. https://doi.org/10.1111/tgis.13053.

Wood, C., B. Sullivan, M. Iliff, D. Fink, and S. Kelling. 2011. "eBird: Engaging Birders in Science and Conservation." *PLOS Biology* 9 (12): e1001220. https://doi.org/10.1371/journal.pbio.1001220.

Xu, J., and Y. Qiang. 2022. "Analysing Information Diffusion in Natural Hazards Using Retweets - a Case Study of 2018 Winter Storm Diego." *Annals of GIS* 28 (2): 213–227. https://doi.org/10.1080/19475683.2021.1954086.

Yan, Y., C. Feng, W. Huang, H. Fan, Y. Wang, and A. Zipf. 2020. "Volunteered Geographic Information Research in the First Decade: A Narrative Review of Selected Journal Articles in GIScience." *International Journal of Geographical Information Science* 34 (9): 1765–1791. https://doi.org/10.1080/13658816.2020.1730848.

Ye, X., and C. Andris. 2021. "Spatial Social Networks in Geographic Information Science." *International Journal of Geographical Information Science* 35 (12): 2375–2379. https://doi.org/10.1080/13658816.2021.2001722.

Zhang, D., Y. Ge, A. Stein, and W. B. Zhang. 2021. "Ranking of VGI Contributor Reputation Using an Evaluation-Based Weighted Pagerank." *Transactions in GIS* 25 (3): 1439–1459. https://doi.org/10.1111/tgis.12735.

Zhang, G. 2021. "Volunteered Geographic Information." In *Geographic Information Science & Technology Body of*

Knowledge, edited by J. P. Wilson. https://doi.org/10.22224/gistbok/2021.1.1.

Zhang, G. 2024. "A Web-Based Geovisualization Framework for Exploratory Analysis of Individual VGI contributor's Participation Characteristics." *Cartography and Geographic Information Science* 1–21. https://doi.org/10.1080/15230406.2023.2293891.

Zhang, G., X. Gong, and D. Zhu. 2024a. "Geographic Proximity and Homophily Effects Drive Social Interactions within VGI Communities: An Example of iNaturalist." *International Journal of Digital Earth* 17 (1): 2297948. https://doi.org/10.1080/17538947.2023.2297948.

Zhang, G., X. Gong, and D. Zhu. 2024b. "Geographic Proximity and Homophily Effects Drive Social Interactions within VGI Communities: An Example of iNaturalist." *International Journal of Digital Earth* 17 (1): 2297948. https://doi.org/10.1080/17538947.2023.2297948.