User Behavior in iNaturalist's City Nature Challenges

iNaturalist

Correlaid Citizen Science project team

Bastian Greshake Tzovaras (*Center for Research and Interdisciplinarity*), Muki Haklay (*University College London*), Katharina Kloppenborg (*Center for Research and Interdisciplinarity*), Florence Lopez (*scieneers GmbH*), Albrecht Mariz (*CorrelAid*), Anna Mogilevskaja, Frank Schlosser (*Humboldt Universität*), Liubov Tupikina (*Center for Research and Interdisciplinarity*), Vadim Voskresenskii (*Weizenbaum Institute for the Networked Society*)

Introduction

These analyses were done in the context of a project call from CorrelAid, a nonprofit network of data scientists who want to volunteer for projects with a good cause. The idea was to analyse the social dynamics of iNaturalist - a citizen science community where users contribute to a open biodiversity database by uploading their nature observations. More specifically, we looked at data from the City Nature Challenges (CNC), annual events where people share sightings recorded in one of the participating cities over a period of some days.

Our research questions revolved around patterns in user behavior and whether there are differences between cities and over time.
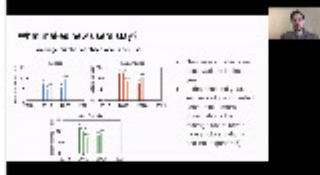
As a basis for our analyses, we used data from iNaturalist's biodiversity database for the CNCs from 2017 to 2020 in three cities - San Francisco, Los Angeles, and London.

We looked into several aspects of the data using various approaches: General statistical data analysis of iNaturalist data separately for the three cities, network analysis of citizen science communities, and spatial aspects including additional data sources available openly (such as green or blue spaces or population).

Our results show...

You can find the code for our results on GitHub.

A presentation of our results is on YouTube:
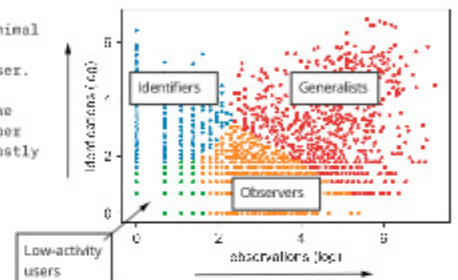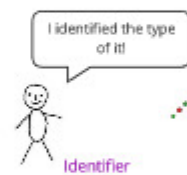

www.youtube.com

Correlaid recording of the project on Citizen science

Citizen science project was started and initiated by studying citizen science data from iNaturalist City Challenge.Project timeline 2020-2021.People involved...

Data Overview

We downloaded data on all **observations uploaded on the iNaturalist platform** during several City Nature Challenges using **their API**.
We obtained data for:

City	Years
London	2018 - 2020
Los Angeles	2017 - 2020
San Francisco	2017 - 2020

The information for each observation we analyzed includes:
The time and location, which user submitted the observation, the identified species, and whether other users have also identified this observations.
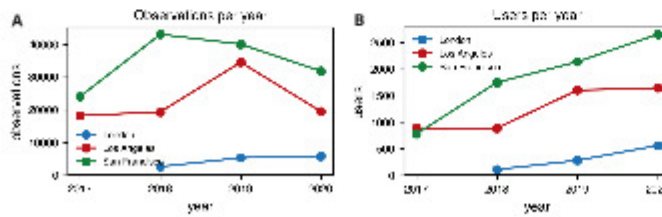
Overall, we find that (A) the **number of users increased** over the years, while (B) the **number of observations roughly stayed the same**. Consequentially, (C) **the average observations per user decreased** over the years.

Attrition of users

An important question for the organization of citizen science events is how new users can be onboarded most successfully. To better understand this, we analyzed: How did onboarding patterns vary between different cities and years? How different are the attrition dynamics of those users who have joined the platform via a challenge and regular users?

New User Onboarding

(A) We analyzed a cohort of **new users**, which didn't participate in 2019 who joined the challenge in 2019. For these new users, we know which users participated in 2020 and were successfully **onboarded** or which didn't and were thus **lost**.

(B) **New users** performed **fewer observations** than established users in **their first year** (2019), but successfully onboarded users **add more observations in the subsequent year** (2020) than the average user.
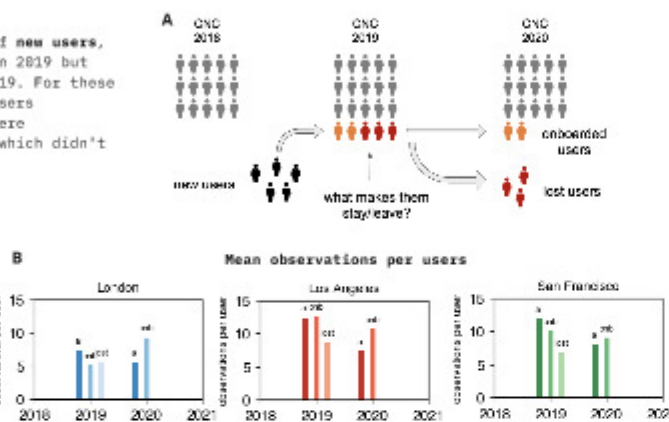
Attrition dynamics of users

The charts show how long the new users stayed on the platform after joining it for 6 months (in San Francisco). X-axis shows the number of months users were active after joining, and Y-axis stands for the percentage of users.

Based on the charts, we can see that regular users stay for longer periods on the platform than challenge participants.

The found difference can be explained by the different intrinsic motivations of these two groups of users. Regular users are interested in the affordances of the platform and ready to stay for longer period if the platform satisfies thir needs, while challenge users, supposedly, rather come on the platform for competing.

That's also interesting that the regular users, who have joined in 2020, stayed for longer periods on the platform than regular users of 2019. One of the potential explanations can be **COVID-19** that limited number of other possible activities for the users.

User Classification

Users can either make an **observation** of a plant, animal or something else, or they can contribute an **identification** of an observation made by another user.

Using the K-Means iterative clustering algorithm, we identify **4 groups of users**, depending on their number of identifications and observations: people who mostly add observations (observers), people who mostly identify (identifiers), those that do both (generalists) and low activity users.
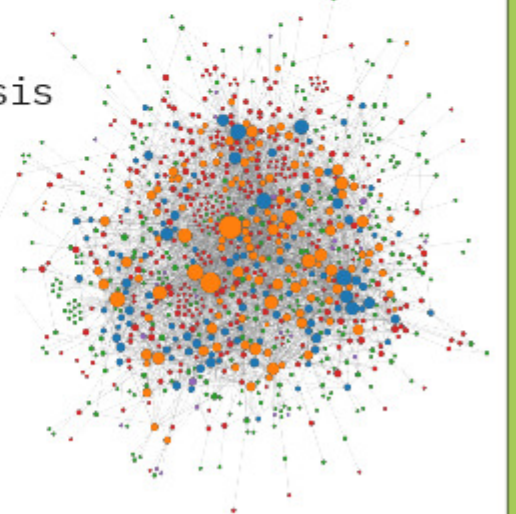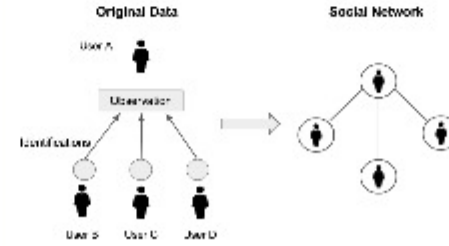

I made photo of a butterfly!

I identified the type of it!

Observer   Identifier

Social network analysis

We extract the **social interaction network** of users by linking those users which identified observations by other users.

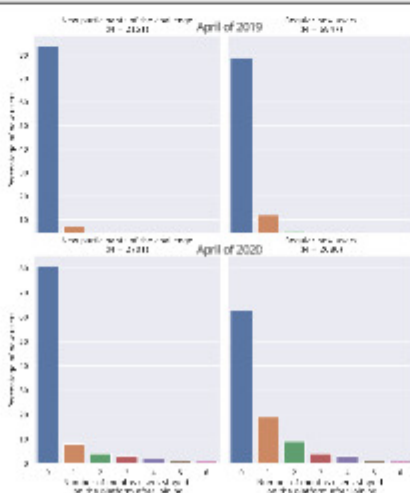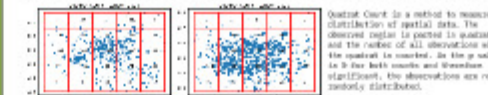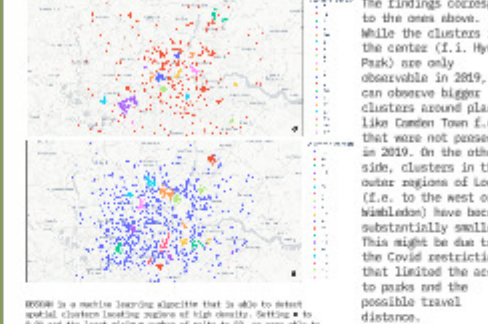We find that **observers are most central in the network**, having interactions with many other users.

Social network for London, 2020

Geospatial analysis

Distribution of observations for the city of London in 2019 and 2020:

We can see that in 2019 there were substantially more observations in the center as well as in the outer regions of London. However, in 2020, there were significantly more observations in the middle circle of the quadrats. Therefore, it seems that observers moved to outer regions within the city but did not travel outside the city to gather data.

Clusters of observations for the city of London in 2019 and 2020:

The findings correspond to the ones above. While the clusters in the center (f.i. Hyde Park) are only observable in 2019, we can observe bigger clusters around places like Camden Town f.e. that were not present in 2019. On the other side, clusters in the outer regions of London (f.e. to the west of Wimbledon) have become substantially smaller. This might be due to the Covid restrictions that limited the access to parks and the possible travel distance.

The interactive maps can be accessed under the link:
https://inaturalist-cnc.correlaid.org/

Greenspaces

Distribution of observations over greenspaces for the city of London:

In 2018 & 2019, the amount of observations made in greenspaces and non-greenspaces is nearly equal to each other. Interestingly, this proportion changes in 2020, meaning that nearly 75% of observations are made in non-greenspaces. This is surely related to **COVID-19**, where parks have been closed in order to minimize social contacts.
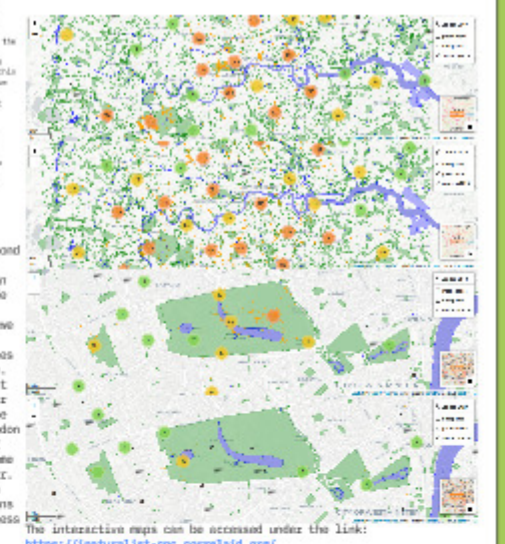
How do the greenspaces in London distribute over the Challenge days?

In 2018 & 2019 the amount of observations made in greenspaces rises significantly on one explicit challenge day, compared to the other days. In the case of 2018 it is known that on the day where greenspace-observations are rising, there was a huge event in the Hyde-Park, leading to the conclusion that this event motivated people to go outside and do more observations in greenspaces. The same may have occurred in 2019. This effect can not be observed in 2020, which may be due to the worldwide COVID-19 pandemic and resulting restrictions and closures of public spaces, such as parks.

Outlook

As an outlook we plan to interact with iNaturalist 2021 and continue collaboration with project participants. It would be interesting to apply current analysis to other citizen science communities.

iNaturalist
iNaturalist is a social network for naturalists. Record your observations of plants and animals, share them with friends and researchers, and learn about the natural world.

CorrelAid
CorrelAid is a non-partisan non-profit network of over 1700 data science enthusiasts who want to change the world through data science. We dedicate our work to the social sector and those organizations that strive for making the world a better place.

miro