A novel application of models of species abundance to better understand OpenStreetMap Community structure and interactions

Peter Mooney
Dept of Computer Science,
Maynooth University,
Ireland

Maynooth University
National University of Ireland Maynooth

Oxford English Dictionary defines Community:

"A body of people or things viewed collectively"

- **A body of people** who live in the same place, usually **sharing a common cultural or ethnic identity**.
- **A group of people** who share the same interests, pursuits, or occupation, esp. when distinct from those of the society in which they live.
- **An online facility**, such as an electronic bulletin board, forum, or chat room, **where users can share information or discuss topics of mutual interest**.
- **A group of animals or plants in the same place**; (Ecology) a group of organisms growing or living together in natural conditions or occupying a specified area.

(Some of) the OSM Community at SotM 2019 in Heidelberg

https://www.flickr.com/photos/thomersch/48769513121/

Learning about our global and local communities in OSM

OpenStreetMap Blog

SUBSCRIBE

via RSS

LANGUAGES

English

BLOGROLL

Survey on global and local communities in OpenStreetMap

Image by Harry Wood. *Background* by Michael Belew (public domain).
OSM logo by Ken Vermette (cc-by-sa & trademarks apply)

Pre-SotM 2019 Survey Response – selected extract from summary

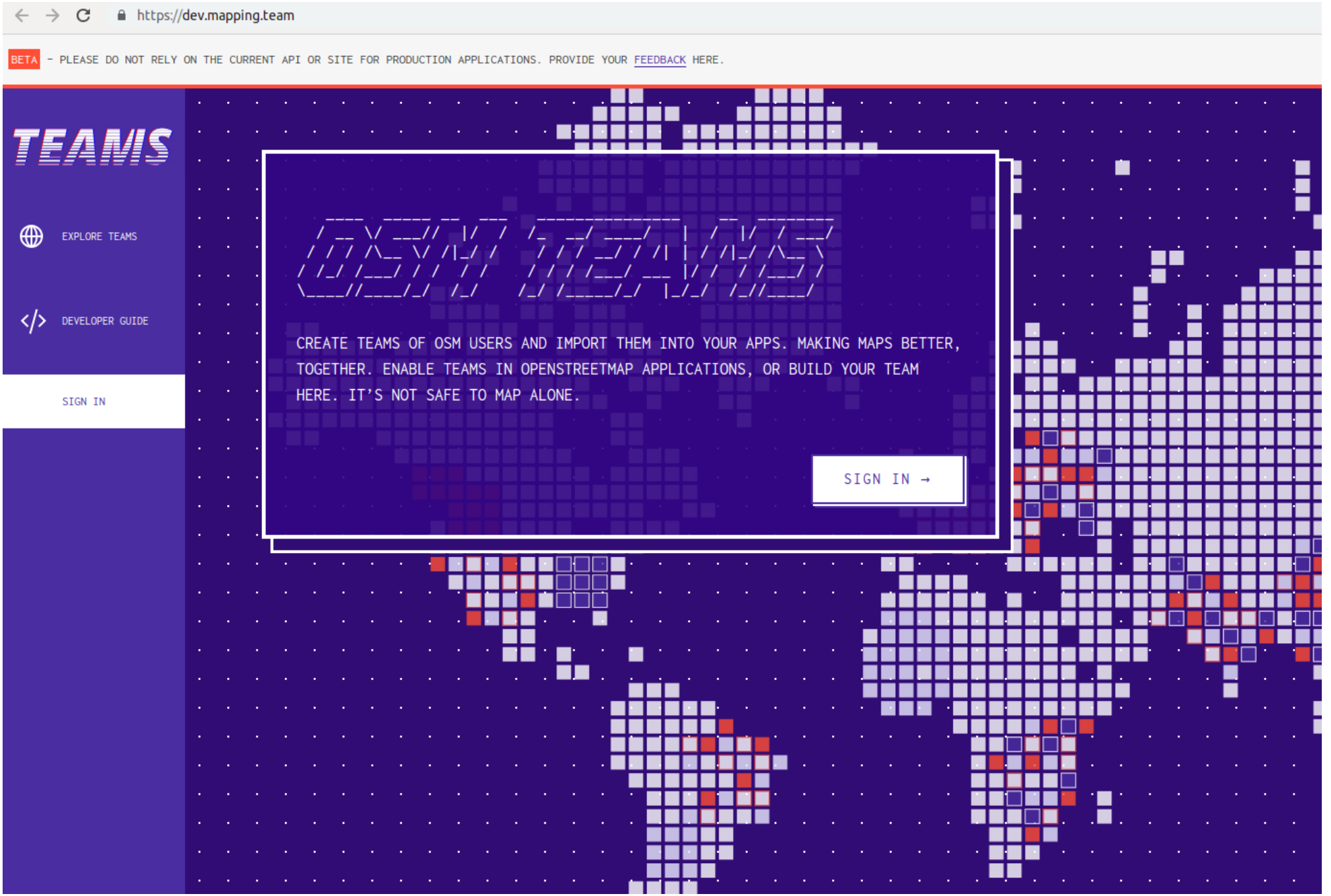
- "**Communities vary** from isolated mappers who only see others editing on the map but don't connect online or offline in person, to places with a full formal organized presence".

- "**The scope of what people consider their community varies** from an individual city or region, to an entire country, to being a part of several places".

- "**We can do more together than apart**. The means to connect and support all these communities is the major challenge".

https://blog.openstreetmap.org/2019/09/17/pre-sotm2019-survey-initial-numbers-and-reflections-from-board-members/

OSM Teams is a recent release



There are many platforms, channels and forums for community information and interaction

https://community.osm.be/resources/europe/germany

DE Deutschland (Germany)

- **Berlin Mailing List** (de)
  This is the mailing list for the Berlin OSM community
- **OpenStreetMap Berlin-Brandenburg Meetup** (de)
  Mappers and OpenStreetMap users in the Berlin area
- **@osmberlin on Telegram** (de)
  OpenStreetMap Berlin Telegram chat
- **OpenStreetMap Berlin Twitter** (de)
  Follow us on Twitter: https://twitter.com/osmberlin
- **OpenStreetMap DE forum** (de)
  OpenStreetMap Germany web forum
- **OpenStreetMap Germany IRC** (de)
  Join #osm-de on irc.oftc.net (port 6667)
- **Talk-de Mailing List** (de)
  Talk-de is the official mailing list for the German OSM community
- **OWL Mailing List** (de)
  This is the mailing list for the Ostwestfalen-Lippe OSM community
- **OpenStreetMap Germany Telegram** (de)
  Join the OpenStreetMap Germany Telegram supergroup at https://t.me/OSM_de
- **OpenStreetMap Germany** (de)
  The platform for information on OpenStreetMap in Germany

From the outside

- A massive global crowd of mappers who continuosly work together to build an open database of the world. These mappers interact via Internet forums, mapping parties, etc

From the inside

- A complex, mutli-layered, dispersed crowd of mappers – a small number of which do a large amount of the work, many others work on their own, interaction happens at specific events.

Academic research has verified many characteristics of the OSM mapper community

- A small proportion of mappers do a very large proportion of the work (*Neis and Zipf, 2012; Caron-Arthur et al, 2014*)
- A very large proportion of mappers only perform a few edits (then possibly leave) (*numerous authors)*
- Some mappers do [a lot of][some] work and then appear to leave the project (*Juhasz et al, 2018*)
- There are different motivations and drivers for mappers (*numerous authors*)
- Some mappers work globally, others locally, others a mixture
- Humanitarian-driven mapping has a major influence on the OSM database
- There appears to be patterns to contributions (day vrs night, event specific, cycles, etc) (*Begin et al, 2018*)

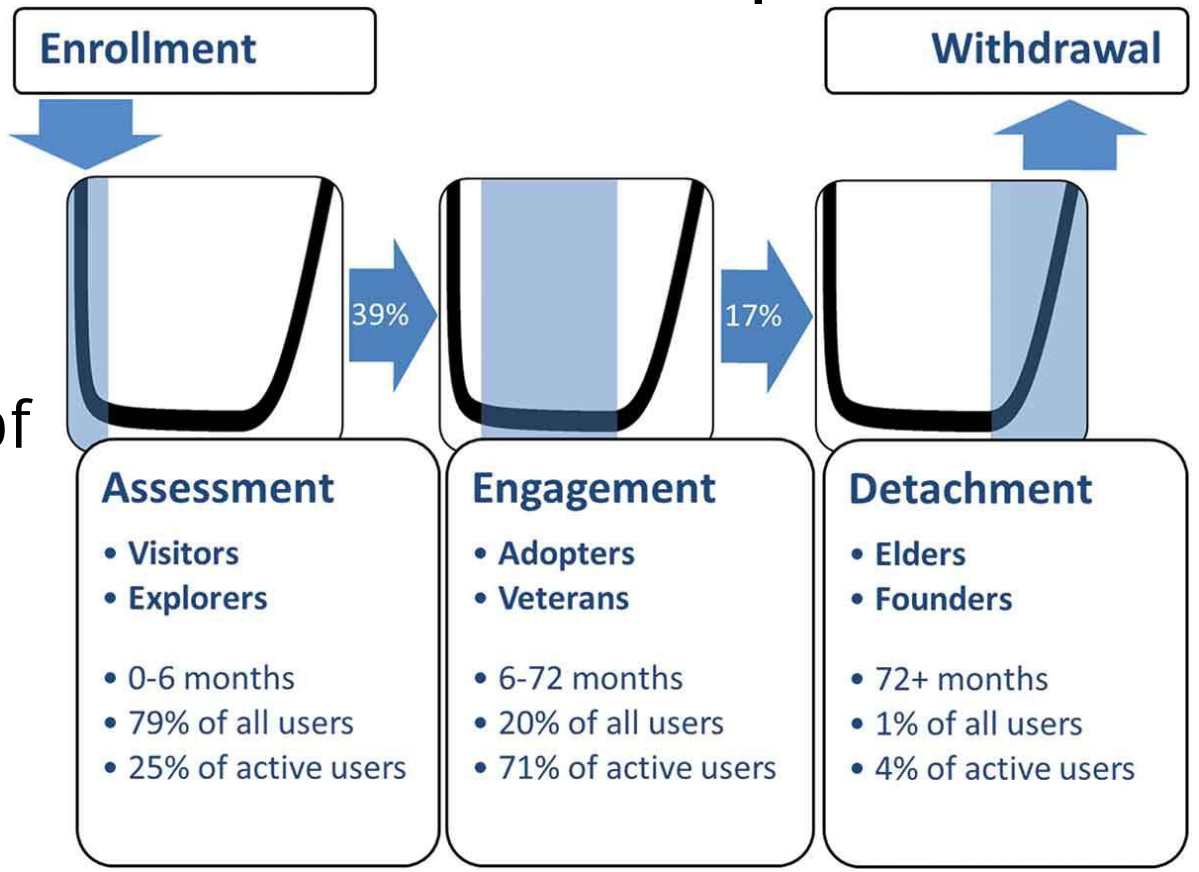
Existing research into OSM Community Structure

- Yang et al (2016) “OSM community has a **vocal minority** (contribute > 95%) and the **silent majority** … contribution inequalities”

https://doi.org/10.3390/ijgi5010005

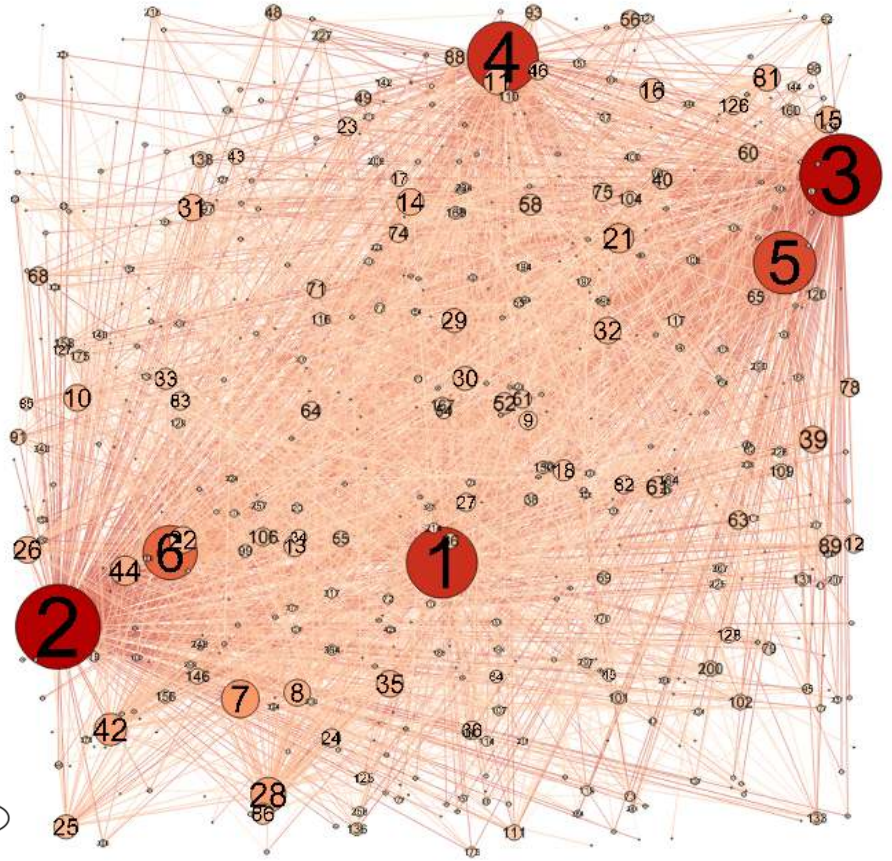
The life cycle of contributors in collaborative online communities –the case of OpenStreetMap

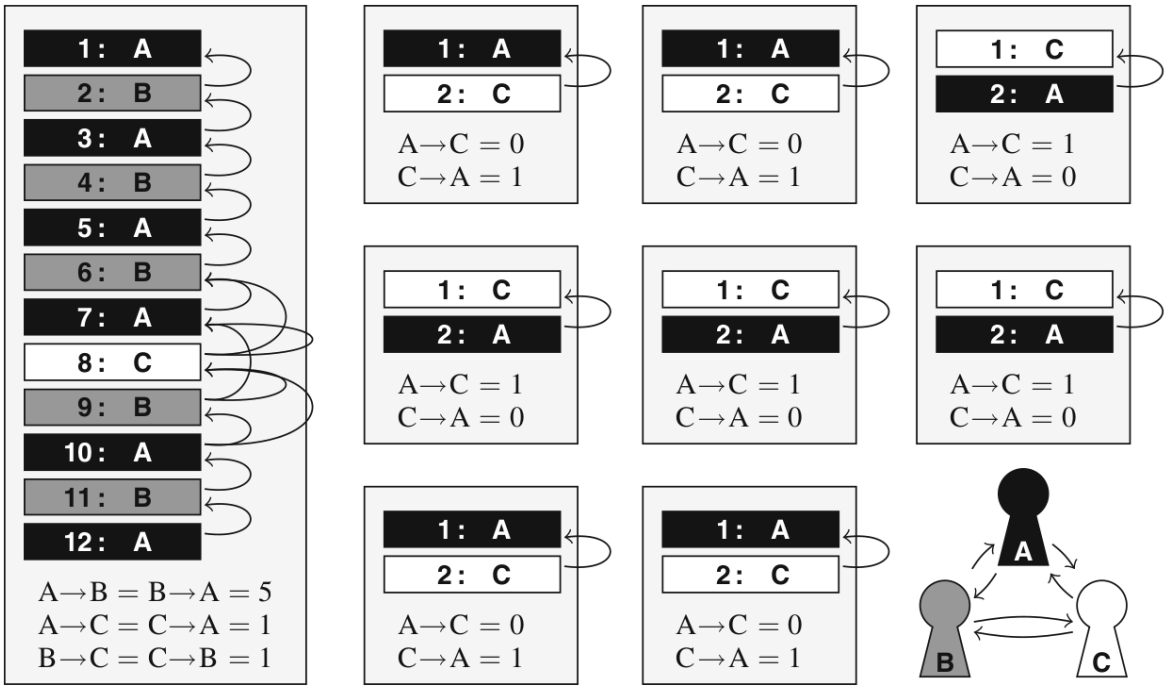
Bégin,Devillers & Roche (2018)



https://doi.org/10.1080/13658816.2018.1458312

Mooney and Corcoran (2013) – Co-editing patterns create a connected graph.
Calculating various metrics shows that **high ranked mappers (workload) interact with the work of many other mappers (high and low rank)**

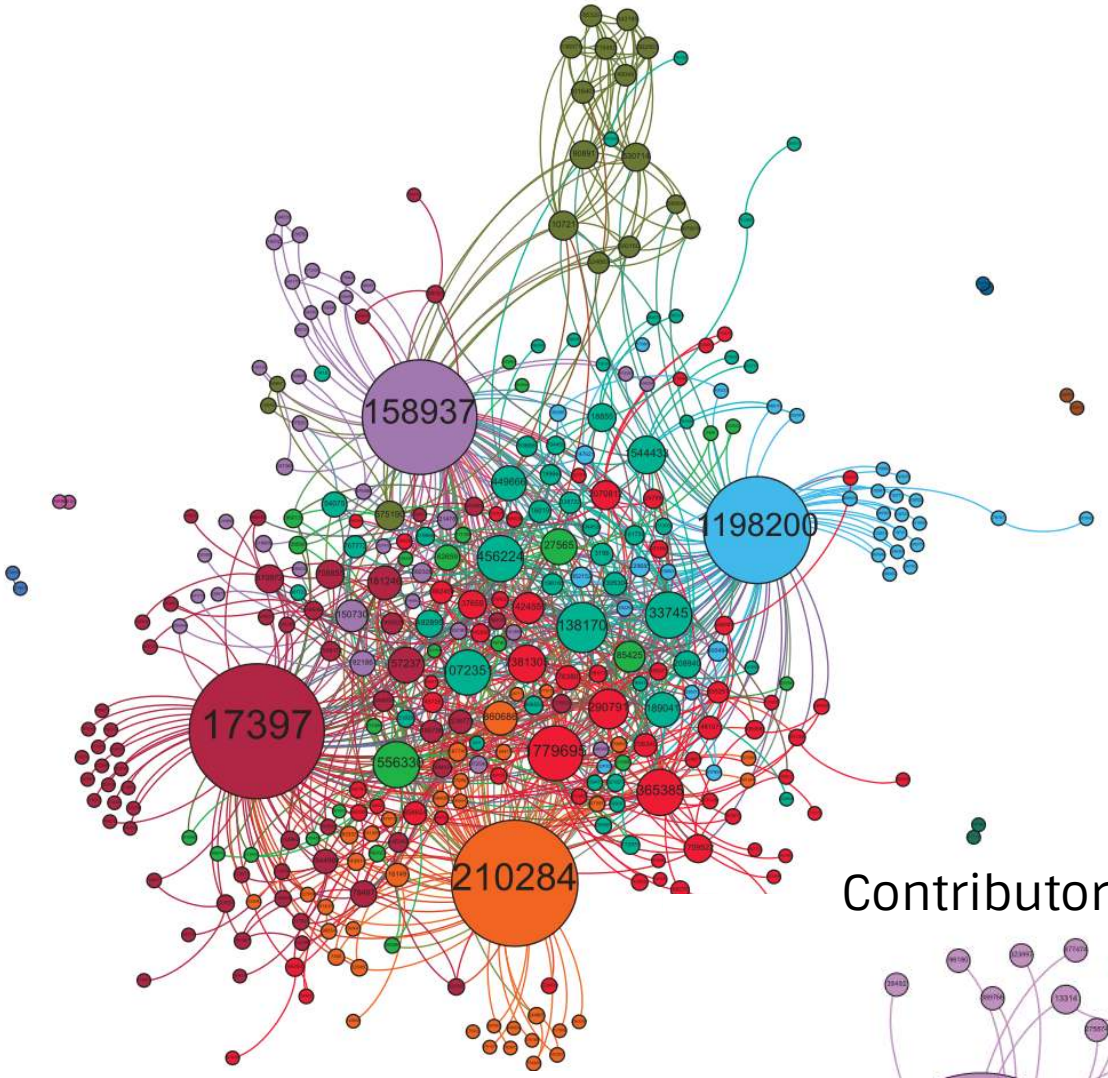




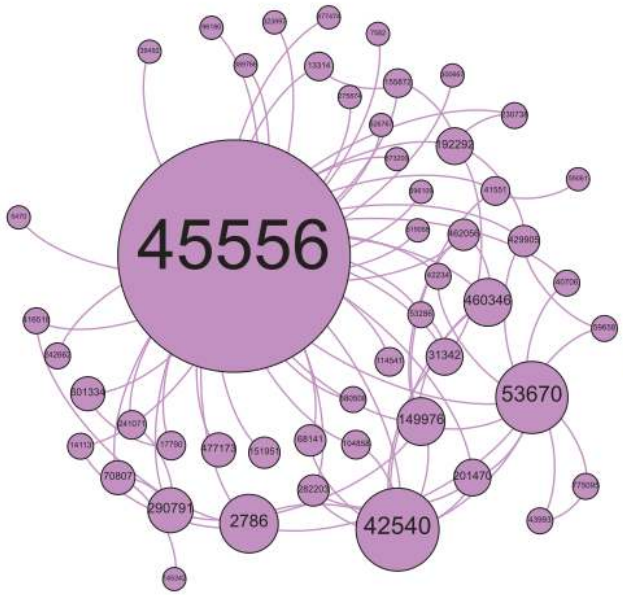
Stein, Kremer, Schleider (2015) – interlocking between map contributors. This allows **quantitative investigation of depth and breath of collaboration** (co-editing)

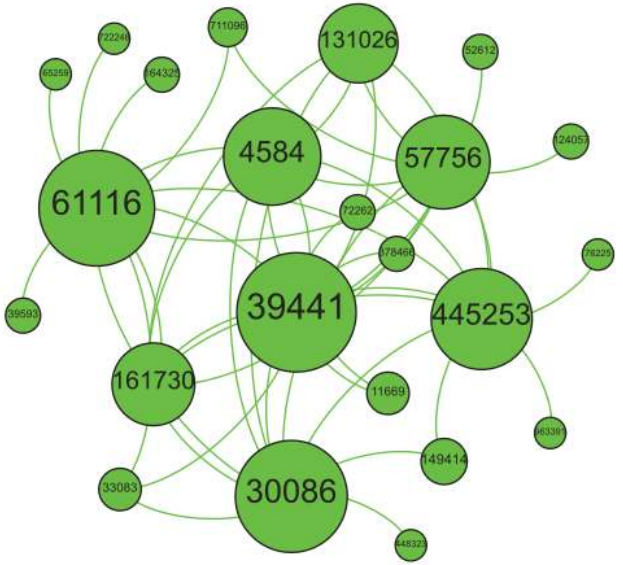
Truong,de Runz, Touya (2018)

Analysis of contributors' behaviour by implementing a multiplex network that gathers various types of collaboration between OSM contributors. Collaboration includes working in the same spatial area. Attempts to identify 'moderator' behaviour



Contributor Centric vrs Unorganised Communities



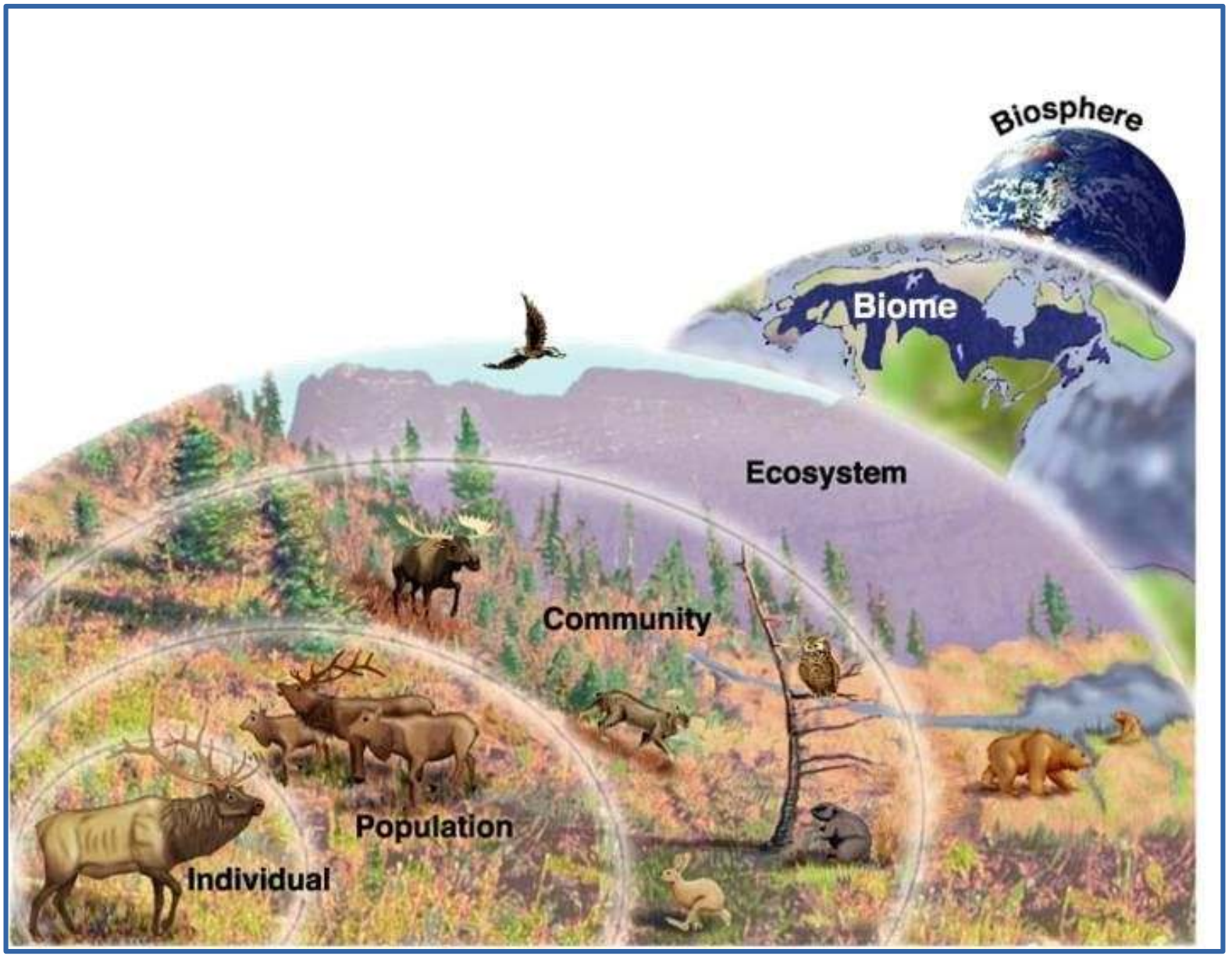


User Ids are annoymous

How are we looking at OSM community structure?

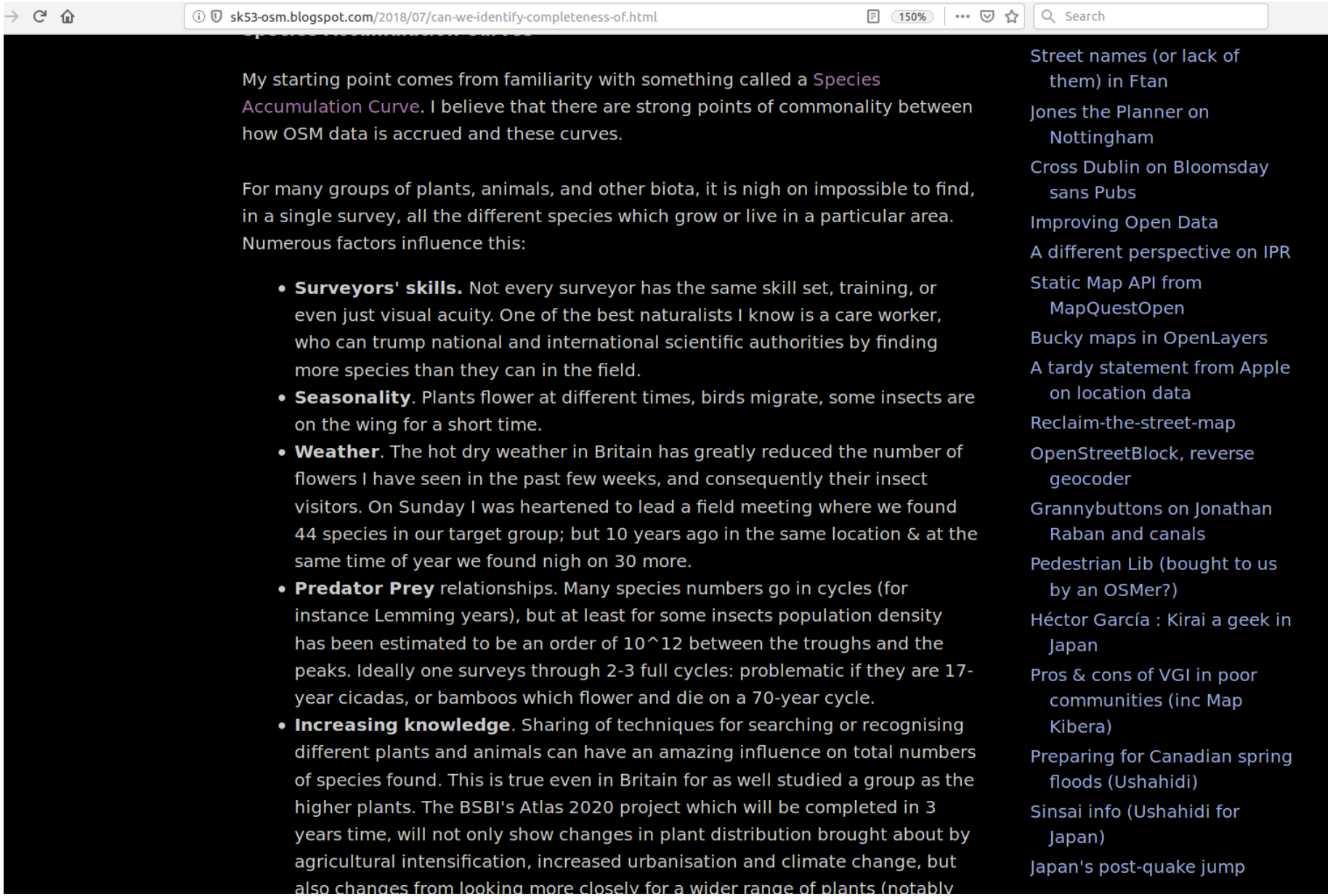
- There is a **focus on quantity/volume of contribution from individuals**
- **Co-editing or editing patterns** (graph theoretic)
- The major focus is on **distinguishing contributors from each other**.
- It is **still difficult to compare different mapping communities**. Mostly we end up saying how complete the OSM database is for a particular community.

A novel idea from Ecology



https://www.pinterest.ie/pin/12420090219916388 3/

Jerry Clough (sk53) linked 'Species Accumulation Curves' to OSM data collection



http://sk53-osm.blogspot.com/2018/07/can-we-identify-completeness-of.html

Water Res. Vol. 18, No. 6, pp. 653–694, 1984
Printed in Great Britain.

0043-1354/84 $3.00 + 0.00

REVIEW

DIVERSITY, BIOTIC AND SIMILARITY INDICES

A REVIEW WITH SPECIAL RELEVANCE TO AQUATIC ECOSYSTEMS

H. G. WASHINGTON
CSIRO Division of Fossil Fuels (formerly Physical Technology Unit), CSIRO Energy and Earth Resources, Ryde, N.S.W. 2112, Australia

(*Received September* 1982)

Abstract—Diversity Indices, Biotic Indices and Similarity Indices are reviewed considering their ecological application, both theoretical and practical. Eighteen diversity indices in eight groups, nineteen biotic indices in ten groups and five similarity indices are examined for their applicability to biological systems, particularly aquatic ecosystems. All the diversity indices were found unsuitable except for Simpsons *D*, Hurlberts PIE, indices based on the theory of Runs (SCI and TU) and possibly McIntosh's *M*. Indices based on information theory, such as *H′* and *H*, although the most commonly used diversity indices were unsatisfactory due to the lack of exploration of their biological relevance. The use of *H′* in aquatic ecosystems is only justified until the more satisfactory indices are adequately field tested.

Biotic indices are highly specialised for a particular type of water pollution, usually organic pollution. Chutter's and Chandler's biotic index appear to be most favoured. In their present form, however, such

similarity indices for water pollution.

The relationship and definitions of diversity and stability are examined. Though discussion is by no means concluded on this topic it is clear that diversity does not automatically lead to either stability or instability but may be found with both. Disturbance to an ecosystem may often increase diversity provided it is of intermediate frequency, yet an ecosystem may also become more diverse without becoming less stable.

The numbers of these three types of indices have become legion, as has their use. It is time that only those indices with biological relevance were used. It is hoped that work will now concentrate on the promising members of the two indices of community structure (diversity and similarity) as these are of broad applicability while biotic indices are limited to one or a few pollutants and one geographical area.

Washington, H. G. 1984. "*Diversity, Biotic and Similarity Indices: A Review with Special Relevance to Aquatic Ecosystems.*" Water Research 18 (6): 653–94. doi:10.1016/0043-1354(84)90164-7.

THEORIES AND MODELS OF SPECIES ABUNDANCE

R. G. HUGHES

School of Biological Sciences, Queen Mary College, University of London, London E1 4NS, United Kingdom

Submitted March 8, 1985; Revised January 21, 1986; Accepted April 17, 1986

It is a common observation that in samples from animal and plant communities most of the individuals belong to a small number of abundant species, whereas most of the species are represented by a small number of individuals. The usual explanations for such patterns of species abundance are founded on one of two mathematical distributions, the log-series and the lognormal, which were initially used to *describe* species abundance in samples (see reviews in May 1975; Pielou 1975). These models may provide a quantitative description of the samples, which is assumed to reflect something of the abundance of the species in the community. However, recent criticisms of the use of these empirical models as descriptors of species abundance (Hughes 1984; Lambshead and Platt 1985) have also undermined their associated explanations, which have been separately criticized for their lack of ecological realism (Pielou 1975; Routledge 1980).

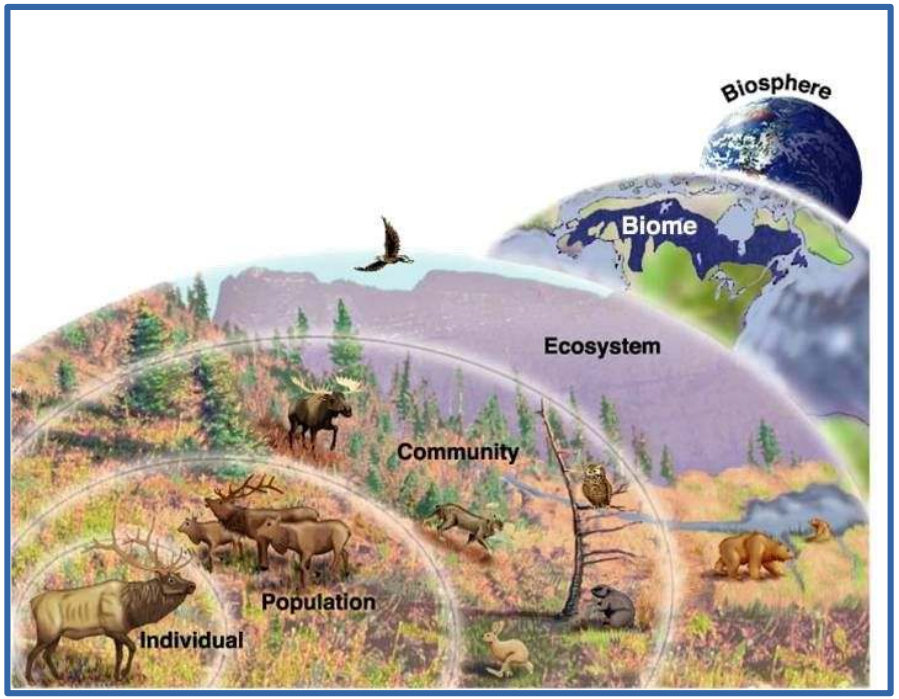
I recently suggested an alternative to these logarithmic models to explain the pattern of species abundance in samples from marine benthic communities

theories of explanation of species abundance, mainly by examining the accuracy of the three models in predicting the pattern of species abundance in a wide range of samples. In addition, the practice of using the log-series and lognormal models as models of community structure, as opposed to models of species abundance in samples, is examined. Other models, notably the gamma and negative binomial distributions, have been used to describe species-abundance patterns, but since they have not been used to provide explanations they contribute little to our ecological understanding and are not considered here.

"The use of indices in ecology has exploded over the last 20 years and a plethora of different approaches and indices has resulted" (Washington, 1984)

- Use a community structure
- **Diversity Indices:** combine data on abundance within species as a single number
- **Similarity Indices:** two samples are compared (on control) – calculate dissimilarity or distance.

How does this ecological idea work?



Ecosystem (where OSM fits in with other crowdsourced geographic information initiatives)

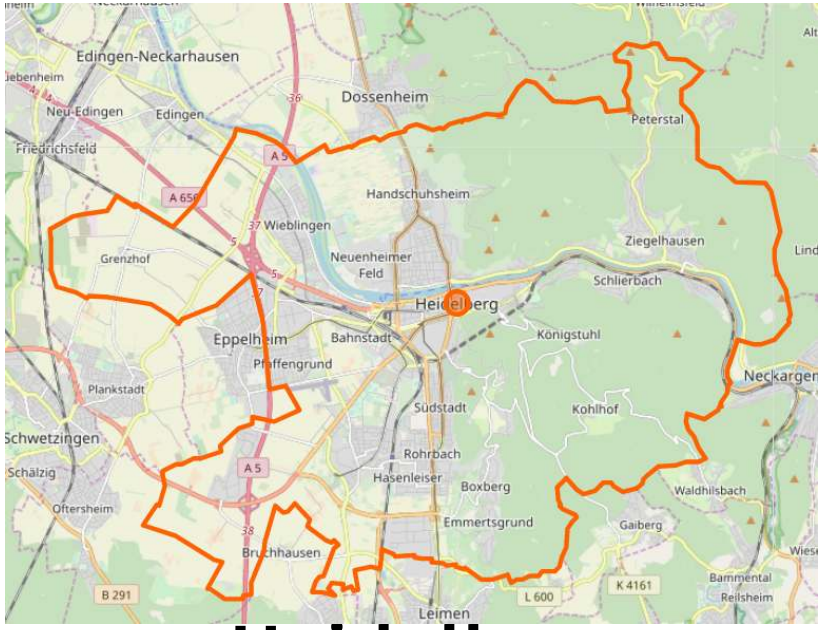
Community (OSM Global, OSM at country level)

Groups of mappers (by country, region, town, meet-up, etc)

Individual Mapper (each contributor to the OSM database)

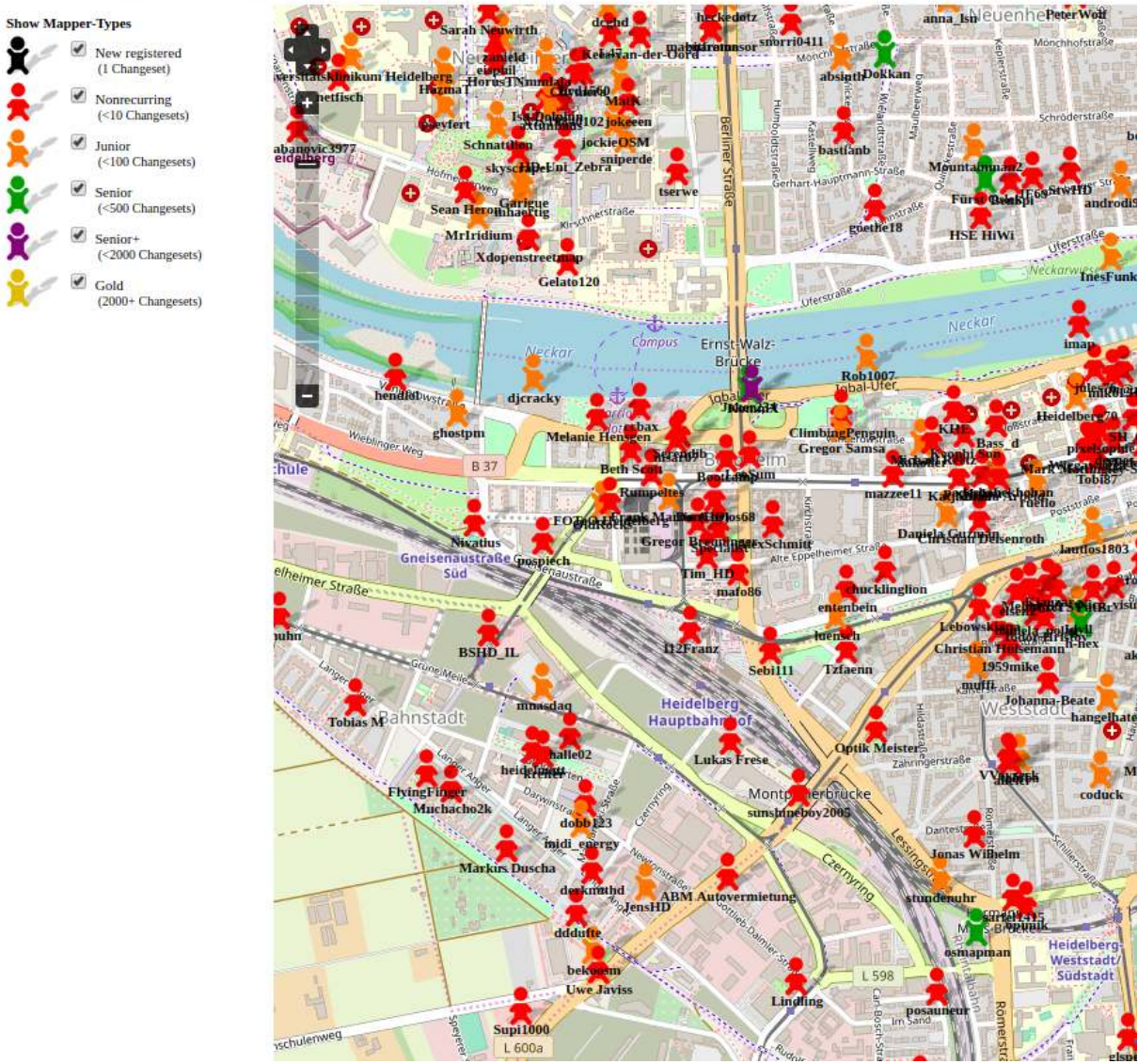
<u>Abundance</u> of particular mapper types (as a single numerical value)

<u>Similarities between two samples of mappers</u> from different communities or ecosystems.



Heidelberg Community

1 Gold (2000+ Changesets)
8 Senior+ (< 2000)
63 Senior (< 500)
515 Junior (< 100)
1508 Non Recurring (< 10)
1675 Newly Registered (1 changeset)



https://resultmaps.neis-one.org/oooc

https://www.openstreetmap.org/relation/285864#map=12/49.4059/8.6836

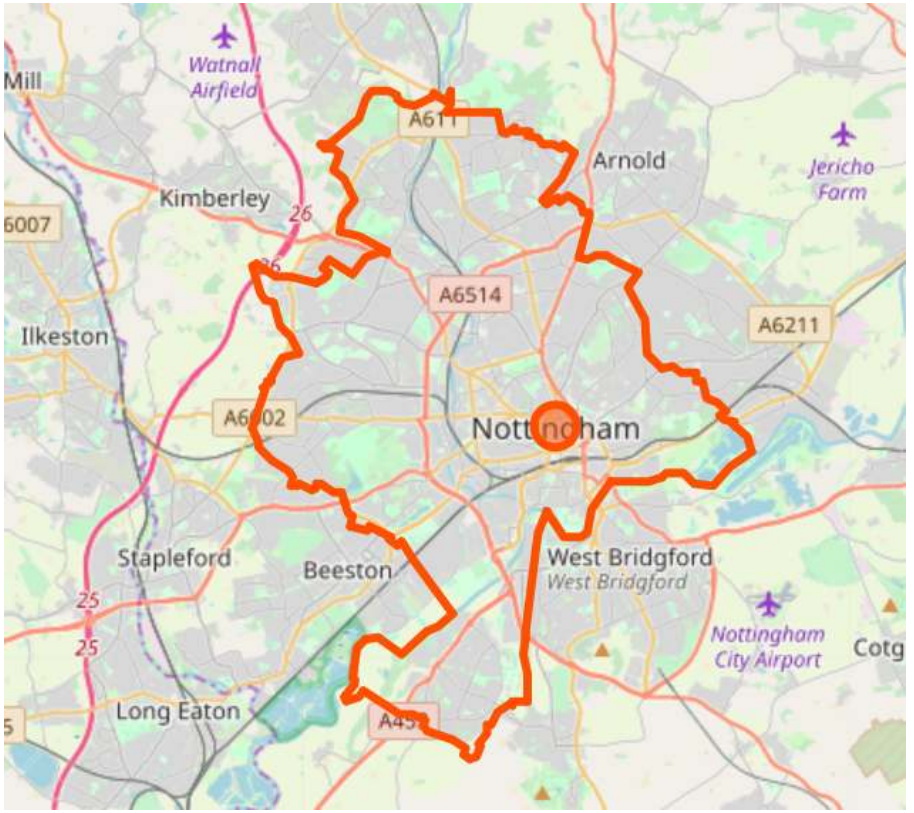
Luxembourg – Community (Country level)

6 Gold (2000+ Changesets)
18 Senior+ (< 2000)
106 Senior (< 500)
749 Junior (< 100)
1878 Non Recurring (< 10)
1941 Newly Registered (1 changeset)

https://osm-internal.download.geofabrik.de/europe/luxembourg.html

Nottingham – Community (City level)

2 Gold (2000+ Changesets)
3 Senior+ (< 2000)
12 Senior (< 500)
143 Junior (< 100)
478 Non Recurring (< 10)
548 Newly Registered (1 changeset)



https://www.openstreetmap.org/relation/123292

$$D = \sum_{i=1}^{S} \frac{n_i(n_i - 1)}{n(n-1)}$$

Simpson's D. The probability that two random individuals will belong to the same species. 0 is infinitely diverse with 1 meaning no diversity.

$$D = \frac{S - 1}{\ln N}$$

Margalef's D (species richness) – linear relation between the number of species and the logarithm of the number of individuals.

$$H = -\sum_{i=1}^{S} \frac{n_i}{n} \ln \frac{ni}{n}$$

Shannons Index: The "magic bullet". The heterogeneity of the population. If two randomly sampled individuals are actually different species. When all species are equally common, then maximum value `ln (S)`.

$$PIE = (\frac{N}{N-1})(1 - \sum_{i=1}^{S} p_i^2)$$

Hurlbert's PIE how each individual in a community can encounter or interact with every other individual in the community. Ranges from 0 uneven to 1 even.

$$M = \frac{n - \sqrt{\sum_{i=1}^{S} n_i^2}}{n - \sqrt{n}}$$

McIntosh's "ecological distance" indicator. ranges from 0 to 1. Baseline 0 where there are no individuals up to the point where every individual is a different species in the community.

Some assumptions and background

- We use the OSM History PBF extracts from GeoFabrik (thanks!)
- We don’t import usernames (just the I.D)
- We only consider the creation and editing of nodes, ways, relations and the tags used
- We assume that OSM mapping communities in different regions share community characteristics

Example 1: Using the Neis Changeset Groups

NEIS Changeset Groups	Nottingham	Heidelberg	Luxemboug
Individuals (Samples)	**1186**	**3770**	**4698**
Species Groups	6	6	6
SimpsonsD *(0..1 no diversity)*	0.45	0.36	0.35
MargalefD *(linear)*	0.71	0.61	0.59
ShannonWeiner *(max 1.79) – all species equally common*	0.94	1.21	1.22
HurlbertsPIE *– encounters (0..1 even)*	0.54	0.63	0.64
McIntoshM *(0... 1 every individual is different)*	0.34	0.41	0.43

Not a very clear picture – no real pattern emerging.

Example 2 – Using Quantiles based on Changeset Production

Quantiles (12) – arbitrary	Nottingham	Heidelberg	Luxemboug
Individuals (Samples)	**1186**	**3770**	**4698**
Species Groups	11	11	11
SimpsonsD *(0..1 no diversity)*	0.62	0.54	0.53
MargalefD *(linear)*	1.41	1.21	1.18
ShannonWeiner *(max 2.39) – all species equally common*	1.24	1.65	1.76
HurlbertsPIE *– encounters (0..1 even)*	0.65	0.72	0.75
McIntoshM *(0... 1 every individual is different)*	0.44	0.52	0.54

Example 3: Specific Mapping Activity

Region	Postboxes and Benches	Mappers (creates and edits)	High volume mappers
Nottingham	790	**52**	**1 x 426 (contributions)**
Heidelberg	2,021	**137**	**1 x 2,020**
Luxembourg	5,836	**306**	**4 x – 1,000**

Example 3: Postboxes and Benches

Groups (4)	Nottingham	Heidelberg	Luxemboug
Individuals (Samples)	**52**	**137**	**306**
Species Groups	4	4	4
SimpsonsD *(0..1 no diversity)*	0.25	0.33	0.43
MargalefD *(linear)*	0.76	0.61	0.52
ShannonWeiner *(max 1.38) – all species equally common*	1.12	1.22	1.15
HurlbertsPIE *– encounters (0..1 even)*	0.72	0.77	0.69
McIntoshM *(0... 1 every individual is different)*	0.34	0.33	0.42

- Excellent – we see that smaller, specific, community, shows that species groups (4) are equally common
- **The individuals within the species groups are reasonably similar** (in terms of their edits)

Future work …

- **More rigorous comparison of communities from different areas** (are communties similar or not…)
- **Opportunities to think about different ways to create species (groups of individuals)** – analyse over different regions, timescales, etc
- **Temporal Changes in Communities** – what changes happen over time?
- **Species development using qualitative** measures (ask mappers what actions they most commonly perform, how do they map?)
- **Use more complex frameworks** for species abundance and diversity such as Hill Numbers (Chao et al, 2014)

Eventually …. I'll make this available on GitHub

`Osmconvert` (command line)

PyOsmium

Python bindings to Osmium Library

LESSON 2019 – Free registration – still available (8th/9th Oct – Zurich)

LESSON 2019: **L**egal **E**thical factor**S** crowd**S**ourced ge**O**graphic i**N**formation

Announcement and 2nd Call for Papers and Participation

1st International Workshop on Legal and Ethical Issues in Crowdsourced Geographic Information

8th - 9th October 2019 (1.5 days)

University of Zurich, Switzerland

Thanks SotM for facilitating the Academic Track

Thanks OpenStreetMap for being the inspiration of our research