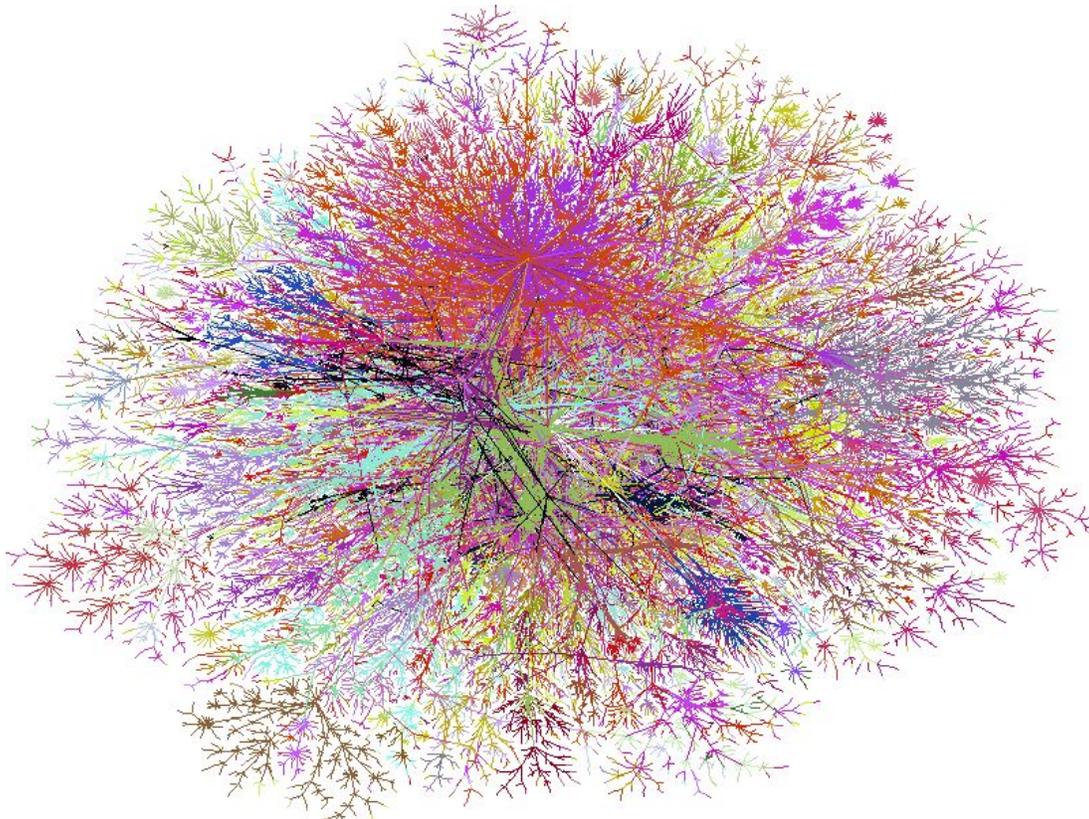# Big Data & Collaborative Economy
State of the art and Data report

DIMM**O**NS
Digital Commons Research Group

Working paper in collaboration with outliers.es
August 2016

Chapter 1: Introduction

Objective

The aim of this chapter is to present the current state of the art in Big Data Analysis applied to collaborative/sharing economy case studies. After some definitions, we will show some academic works, some white papers and reports, some relevant data visualizations; and finally we will wrap-up with some conclusions and opportunities in the field.

What is collaborative economy Big Data Analysis?

In our understanding, Big Data Analysis applied to Collaborative Economy, must comply with the following features:

1. The analysis should be based upon datasets derived from user data (i.e: activity of the users inside the internet-backed system from which the particular service is provided) or aggregations based of that user-level data (e.g: system-wide transactions per time unit)
2. The amount of data must be big enough to be considered Big Data. We are not looking here for survey-based studies, for instance. Big Data has a very elusive definition. Here we considered that if data comes from user data or their pertinent aggregation, we're dealing with Big Data

Collaborative consumption vs Commons Collaborative Economy

A distinction must be made between two different kind of services provided into the more broad 'umbrella' of 'collaborative economy' or 'sharing economy'. As of today, a shared definition for these terms is lacking[1].

We propose a distinction based on the commercial or community orientation of the services, and thus:

- On the one hand, we have Commons Collaborative Economy services and platforms, such as CouchSurfing (which has a strong community orientation)
- On the other, we have classic collaborative consumption services and platforms, such as AirBnB (which has a clear commercial orientation)

We are talking about 'Commons Collaborative Economy' services or platforms when:

1. The sense of community is strong
2. Usually, the provider of the service does not charge any commission or fee for each transaction (and for each end of the transaction)
3. The idea is to share idle resources between pairs or communities

We are talking about 'Collaborative Consumption' services or platforms when:

---

[1]

http://www.collaborativeconsumption.com/2013/11/22/the-sharing-economy-lacks-a-shared-definition/

1. There's a clear commercial orientation (i.e: money first, community second)
2. Usually the provider of the service does charge commissions or fees for each transaction or user registered
3. The idea is to provide a cheaper or more efficient alternatives to other 'non-collaborative' services.

Examples of 'Commons Collaborative Economy' services can be: Guifi.net, Couchsurfing or Fairmondo
Examples of 'Collaborative Consumption' consumption services can be: Uber or AirBnB

For a further distinction between Couchsurfing and AirBnB and their underlying models (from a user perspective) please visit
https://www.tripping.com/industry/rental-companies/airbnb-vs-couchsurfing

For an in-depth description of 'Common Collaborative Economies' please visit
http://procomuns.net/wp-content/uploads/2016/05/CommonsDeclarationPolicies_eng_v03.pdf

Data analysis main aims

In a broad sense, in the field of Data Science, analysis are conducted in order to satisfy some of these two needs:
1. Prove a hypothesis well known beforehand. For example: You have a strong intuition about the correlations between two variables and want to confirm that with data analysis techniques.
2. Know your data: what are the questions that can be answered by a particular dataset? What are the structures and relationships embedded in this particular dataset?

Note that Data Visualization is paramount in the second point but also have a non-trivial role in the first one.

Ultimately, one have to 'come out' of a Data Analysis cycle with (a) new insights and/or (b) new questions for further explorations.

Other classifications

In our case ('collaborative economy'), as will be seen below, the 'hot topics' amongst academic material and white papers as of today, are the following:

1. Rate/Characteristics/motivations of adoption of these new human interchange models (Adoption)
2. Implications to local or global economies of some services/platforms (Impact)

3. Behavioral, Geographical and physical models that help us understand the dynamics underneath these services/platforms (Modelling)
4. Legal or ethical issues related to some services/platforms (Ethics)

Finally, we have searched for a shared taxonomy to classify these new models, looking for a compromise between extension and coverage. Probably the most practical we have found, located at http://www.slideshare.net/jeremiah_owyang/sharingnewbuying, divides the Collaborative Economy ecosystem into four broad categories: Goods, Services, Transportation, Space and Money, each one of these with sub-categories depicted below

Our goal in this introduction is to present reference material in a well-balanced manner, i.e: find works in both the academic and commercial fields, but also a nice mix of works related to both the 'Commons Collaborative Economy' vs 'Collaborative consumption', a nice balance between all the hot topics presented earlier (Adoption, Impact, Modelling, Ethics), and a good coverage of all the 'categories' above ('Goods', 'Services', 'Transportation', 'Space', 'Money)

Academic work in the field of collaborative economy data analysis

We present below a series of academic papers and studies that we consider relevant to the state of the art.

Literature Review of Couchsurfing Research

Topics: Modelling and Adoption
Alexander Ronzhyn, March 24, 2016. Source: http://ronzhyn.com/files/cs_review.pdf

A really good review and compendium of all the academic literature published about Couchsurfing (totalling 24 references as of 12th August 2016). We find papers around topics like: ratings modelling and design (Modelling), Motivations (Adoption), Homophily (Modelling), Language analysis of reviews (Modelling), Trust analysis (Modelling)

Not all the papers use Big Data analysis, but here are some selected entries that do so:

"Rating Friends Without Making Enemies"
Lada A. Adamic, Debra Lauterbach, Chun-Yuen Teng, Mark S. Ackerman
University of Michigan
https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/viewFile/2837/3280

Big Data analysis of reviews, plus interviews and surveys in order to quantitatively differentiate between ratings of trust and friendship

"A Case for Space: Physical and Virtual Location Requirements for Internet-based Social Networks"
Edward Pultar, Martin Raubal
University of California
http://idwebhost-202-147.ethz.ch/Publications/RefConferences/PultarRaubalLBSN2009_v5.pdf

Big Data analysis of cities, activities and people (communities) inside CouchSurfing, modelling the characteristics of the communities as both physical and virtual. Based on site-scraping techniques.

"Conveying the Message of Trust through Written Texts in CouchSurfing.org"
Alexander Ronzhyn and Eugenia Kuznetsova

http://synthesis.enl.uoa.gr/journal/alexander-ronzhyn-and-eugenia-kuznetsova.html

Natural Language Processing techniques attempting to model "the language of trust", in terms part-of-speech analysis, bag-of-words frequencies and metaphors used. Based on profile and references sampling from major cities.

Car Related Platforms

An Analysis of the Labor Market for Uber's Driver-Partners in the United States

Jonathan V. Hall (Uber), Alan B. Krueger (Princeton)
2015
https://s3.amazonaws.com/uber-static/comms/PDF/Uber_Driver-Partners_Hall_Kreuger_2015.pdf

An analysis of the demographics involved in Uber drivers, their evolution and their working patterns. Based on aggregated data from Uber.

Income Targeting and the Ridesharing Market

Michael Sheldon
2016
http://static1.squarespace.com/static/56500157e4b0cb706005352d/t/56da1114e707ebbe8e963ffc/1457131797556/IncomeTargetingFeb16.pdf

Study of income elasticity based on different factors. Trip and supply activity data from undisclosed peer-to-peer ride sharing platform

Quantifying the benefits of vehicle pooling with shareability networks

Paolo Santi et al.
Massachusetts Institute of Technology, Istituto di Informatica e Telematica del CNR, Cornell University, Ithaca.
https://arxiv.org/pdf/1310.2963.pdf

Based on a study of 150 million taxi trips in New York, the study tries to build a model for targeting the benefits of car-sharing under different circumstances and locations.

Peeking Beneath the Hood of Uber

Le Chen et al.

Northeastern University
2015
http://conferences.sigcomm.org/imc/2015/papers/p495.pdf

Uber Surge price algorithm analysis and its consequences. Based on Uber app emulation data retrieval.

AirBnB

Who Benefits from the "Sharing" Economy of Airbnb?

Giovanni Quattrone et al.
University College London, Boston University, Bell Laboratories
2016
http://researchswinger.org/publications/quattrone16airbnb.pdf

AirBnB presence in different socio-economic level areas, and how to regulate its impact. Based on data scraped from AirBnB.

The Rise of the Sharing Economy: Estimating the Impact of Airbnb on the Hotel Industry

Georgios Zervas et al.
Boston University, University of Southern California
2013
http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2366898

Analysis of the (negative) impact of AirBnB in the hotel industry and its spatial distribution. From AirBnB listings scraping + internal hotel data.

Market Platforms

Engineering Trust

Gary Bolton et al.
Pennsylvania State University, University of New South Wales, University of Cologne,
2012
http://ben.orsee.org/papers/engineering_trust.pdf

User feedback analysis optimization in reputation information. Based on data scraped from Ebay and other platforms

Proceedings, reports and white papers

We've found several case studies that, although not presented as papers, we consider relevant because of their scope, methods or insights

FiveThirtyEight series on Uber Data

http://fivethirtyeight.com/features/uber-is-serving-new-yorks-outer-boroughs-more-than-taxis-are/
http://fivethirtyeight.com/features/public-transit-should-be-ubers-new-best-friend/
http://fivethirtyeight.com/features/uber-is-taking-millions-of-manhattan-rides-away-from-taxis/
http://fivethirtyeight.com/features/is-uber-making-nyc-rush-hour-traffic-worse/

Series of articles from http://fivethirtyeight.com/ dissecting the behaviour of Uber vehicles and market in New York (impact and modelling). Includes infographics, maps and statistical graphics. Very relevant in terms of their final findings or insights. Data is from Uber and open here: https://github.com/fivethirtyeight/uber-tlc-foil-response

Uber and the economic impact of sharing economy platforms (Bruegel)
http://bruegel.org/2016/02/uber-and-the-economic-impact-of-sharing-economy-platforms/

Characterization of Uber drivers and the impact Uber is having on the taxi market in several cities. Data aggregated from Uber. Relevant because of the cross-sectional analysis and whole scope.

Analysis of the impact of short-term rentals on housing (City and county of San Francisco)

http://www.sfbos.org/Modules/ShowDocument.aspx?documentid=52601
Analysis of AirBnB listings in San Francisco and possible policies to differentiate between occasional renting (legal) and commercial renting (irregular) inside AirBnB platform. This was the first model built upon irregular postings in AirBnB and the first attempt to characterize AirBnB ethical status in a collaborative context.

Report: Office of the Attorney General of the State of New York - Airbnb in the city (2014)

Access: http://www.ag.ny.gov/pdfs/Airbnb%20report.pdf

Methodology: Data analysis: On May 14, 2014, **NYAG served Airbnb with a subpoena for detailed information about rental transactions on its platform** . Shortly thereafter, and pursuant to an agreement dated May 20, 2014 , Airbnb shared data with NYAG reflecting certain rental transactions in an anonymized format (the "Data") .

Data visualization-centered works

Inside AirBnB

http://insideairbnb.com/san-francisco/
A nice site offering datasets gathered by scraping Airbnb (available for a lot of cities). You can also visualize the data into maps and dig into several variables on-the-fly

"Uber around the world"
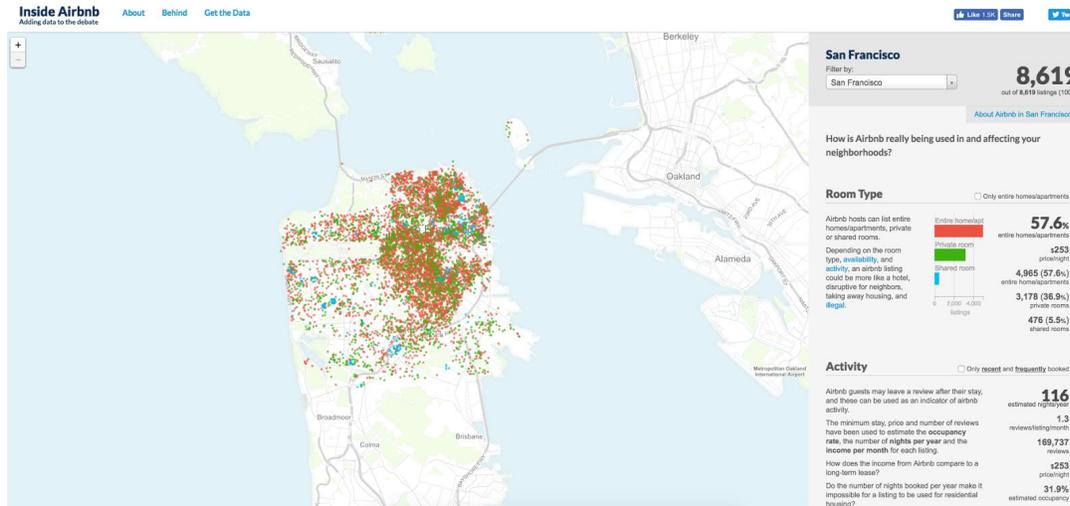
http://flowingcity.com/visualization/uber-around-the-world/
Data visualizations (ala Eric Fischer) displaying Uber trips density in several cities

"Sharing Economy & Italy and the Sharing Economy" - Data Visualization

http://labs.densitydesign.org/dd11/es1/g1_sharing-economy/index.html
Interactive infographics about the sharing economy (impact, market and users), both globally
and in Italy

Conclusions

Common practices in collaborative economy data analysis

As seen in the previous paragraphs, there are some recurrent themes in the works presented:

1. The vast majority of works, acquired the data via web-scraping (Airbnb, Ebay, CouchSurfing), while Uber-related work is normally given from Uber itself. This paradox still need to be resolved. The collaborative economy still lacks their 'Twitter' equivalent (flexible API with really good coverage of data)
2. The most frequent topics include Impact (Airbnb), Model & Trust Dynamics (Ebay, CouchSurfing), and both Impact and economics modelling (Uber)
3. The methods employed to analyze the data vary widely: Natural Language Processing, Statistical Methods (Correlations and Factorial Analysis), Geospatial analysis and Machine Learning (Clustering and Classification

Aside from that, in general, the methodologies employed do not differ significantly from those employed by Big Data research and Data Science applied to other field

Some conclusions

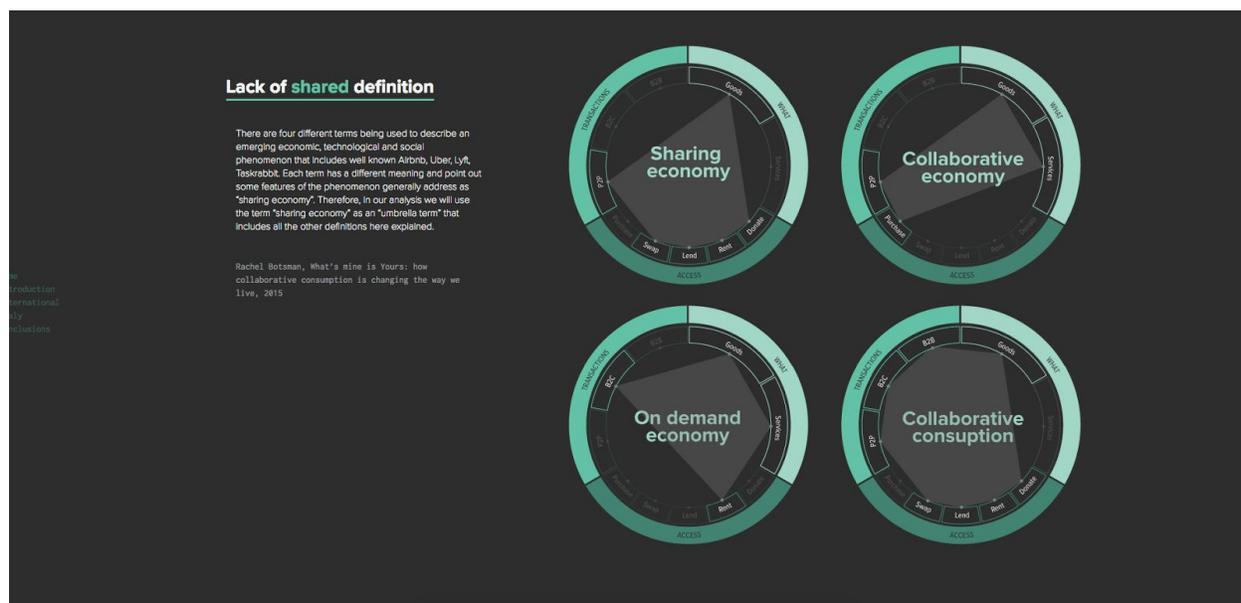1. The collaborative economy still lacks their 'Twitter' equivalent (flexible API with really good coverage of data)
2. There is a general concern of the socio-economical impact of these collaborative consumption platforms, both in terms of the local market and in the labour force (Uber and Airbnb)
3. There is another general concern about the algorithms that automatically drive the dynamics of the service. Trust dynamics in both Ebay and Couchsurfing and surge prices in the case of Uber

Opportunities

Reading into the conclusions and common practices and topics, there is a huge opportunity in the fact that to really understand and model these platforms, you have to cross it with descriptive data of very different nature and origin, such as City Council Data, local market data, human transit data and Open Data, which can put to use lots of these datasets and provide novel insights due to the intermingling of different kinds of data.

On the other hand, the analysis and modelling of the dynamics (both human and algorithmically driven) implied in these human interchanges can give us really powerful and unforeseen insights into community and peer to peer market dynamic, which can really advance the state of the art of Data Science and Big Data Analysis.

Chapter 2: Indirect Data

Objective

Definition of possible strategies aimed at building or combining different indicators related to the collaborative economy, starting from available 'indirect' data such as Wikidata, Twitter, Facebook, Public Data and so on). Some of this strategies should also be oriented towards studying collaborative economy in a 'city context', its impact and related public policies implemented

As a first step, every relevant data source will be listed. Afterwards, a section is included where some strategies are proposed for (a) creating derived indicators and (b) how these indicators can be combined.

Data Sources. Listing and characterization

Three categories have been used to classify indicator formulation:

1. **Digital impact:** Any data that can help us analyze the impact of all the different collaborative economy services in the population (through the analysis of 'digital channels')
2. **Market:** If the data source help us implementing a time-series analysis of different metrics related to the economy of raw materials/technologies involved in the services, in order to monitor and track 'traditional services' markets vs those that are collaborative-economy based
3. **Use:** Data sources related directly with 'client apps' involved in the service provided
4. **City impact:** Datasets / Listings of datasets about the direct impact of specific collaborative-economy services in different cities (or that can help us infer the aforementioned impact)

Digital Impact

Wikipedia and related data

Wikipedia edit-stream

https://wikitech.wikimedia.org/wiki/Stream.wikimedia.org

Allow us to 'hear' the live 'streaming' of Wikipedia edits. Geographical origin of the each edit can also be tracked. It can be useful to have a notion of the activity and geographical segmentation of different (related) articles editors (e.g: https://en.wikipedia.org/wiki/Uber_(company), https://en.wikipedia.org/wiki/Airbnb, https://en.wikipedia.org/wiki/CouchSurfing)

Controversy measurements

Following a similar strategy such as the one used in the article http://airwiki.ws.dei.polimi.it/index.php/When_the_Wikipedians_Talk, a 'controversy scale' can be implemented (based on 'reply' chain in article-discussion pages). Moreover, this 'controversy scale' can be tracked through time and follow its time evolution, in order to detect 'hot' articles (in the controversy sense) and specific interesting moments in time

Pageviews per article

Using the service https://en.wikipedia.org/wiki/Wikipedia:Pageview_statistics, we can get access stats for every Wikipedia article (e.g: https://tools.wmflabs.org/pageviews/?project=en.wikipedia.org&platform=all-access&agent=user &range=latest-20&pages=Airbnb), also with language segmentation information.

Trends

The idea is to track the 'digital interest' of the different relevant services, with the added possibility of geographic segmentation.

Google Trends

Trends search (search volume) for different terms, with the added possibility of analyzing the term timeline and geographical distributions (including the country from which the queries are originated)

Some examples:
- 'News' trends about 'sharing economy' and 'collaborative economy' (https://www.google.com/trends/explore?cat=16&date=all&q=sharing%20economy,collaborative%20economy)
- 'Digital commons' vs 'collaborative economy' https://www.google.com/trends/explore?cat=16&date=all&q=sharing%20economy,collaborative%20economy,digital%20commons
- 'Web search' trends of 'carsharing' vs 'homesharing' vs 'crowdfounding' (https://www.google.com/trends/explore?date=all&q=%2Fm%2F03t80x,home%20sharing,%2Fm%2F05ztt99)
- Trend timelines: 'crisis' vs 'sharing economy' vs 'collaborative economy' vs 'airbnb' vs 'uber' (https://www.google.com/trends/explore?date=all&q=crisis,sharing%20economy,collaborative%20economy,%2Fm%2F0gx0wlr,%2Fm%2F0svqyn7)

Twitter API 'listening'

Based on Twitter Streaming API (https://dev.twitter.com/streaming/overview), a number of specific accounts can be 'followed' (e.g: those owned by the different services) or a number of specific and relevant keywords'/ 'hashtags' can be followed too.

Based on this information, a number of strategies can be implemented:

- User communities can be analyzed (those that are involved in interaction with specific services), in order to get to know their network 'structure':
- Twitter API user-location or tweet geolocation can be used to analyze the geographic distribution of some 'hot' topics, hashtags or accounts

- Finally, 'listening' to a Twitter account stream of 'replies' and through quantitative techniques or even qualitative techniques (language analysis), we can zoom the lens into specific dialogues between a service account and its users and get to know how many of them are 'complaints' or 'congratulations/thanks' about the brand or service

Market

Cost of living and assets related to collaborative economy

Numbeo

Database about the economics and costs (via 'crowdsourcing' of transport, lodging, finances and food (also about average household incomes), segmented by country and city. http://www.numbeo.com/cost-of-living/country_result.jsp?country=Spain

Relevant examples:

- Costs for different 'domains' in which the sharing economy operated (average flat rent, taxi fares and so on)
- Average monthly income in different cities
- Cost of living spread into different metrics (from 2009 until mid-2016) (http://www.numbeo.com/cost-of-living/region_rankings.jsp?title=2016-mid&region=150, http://www.numbeo.com/cost-of-living/region_rankings.jsp?title=2009&region=150)

Asset cost estimation

Ways to estimate the unit cost (or time cost) of implied assets (real state, vehicles and so on..) can be found with the help of different information services.

Some examples:

- Uber fares in different countries / cities (http://uberestimate.com/prices/)
- Average listing price in Airbnb (http://insideairbnb.com/)
- Lyft fares depending on trip origin and destination (https://www.lyft.com/developers)

Use

Apps

Number of approximate app download, with historic and geographical analysis included (previous registration is needed)
https://www.appannie.com/apps/google-play/app/com.airbnb.android/details/

Service activity

Through some datasets included in the present study or with the help of web scraping techniques, a global notion of 'real' service activity can be reached (besides 'client app' downloads)

Some examples:

- Thanks to periodic downloads from http://insideairbnb.com/ we can know the number of available properties/rooms ('listings') in Airbnb. With the additional information (included) of availability and calendar of every listing, we can also know the exact daily offer as well as the prices.
- Using Zipcar API (https://developer.zipcar.com/documentation/location), we can get to know the vehicle 'fleet' near a specific location. Using this feature, we can extrapolate our findings to a whole city.
- Also, using Lyft API (https://www.lyft.com/developers), the vehicle 'fleet' for each city can be determined.

City impact

As of August 2016, we have not found specific 'datasets' which could be used to measure collaborative economy direct impact on cities. We consider the following sources relevant, though:

List of legal actions and protests against Uber in different countries and cities
https://en.wikipedia.org/wiki/Uber_protests_and_legal_actions

Law modifications in different California cities regarding Airbnb impact
http://www.losangeleslandlordlawyer.com/california-airbnb-laws/

On the other hand:

If we want to help in the analysis of the impact of 'casual' tourism households in Spanish, the number of hotel beds in this 'INE' link can be helpful:
http://www.ine.es/jaxiT3/Tabla.htm?t=2076

'Car sharing' and 'Car riding' impact:
We have not found accurate statistics of taxi 'fleet' in cities, but this Wikipedia link has approximate numbers for different countries and cities:
https://es.wikipedia.org/wiki/Anexo:Taxis_en_el_mundo#Espa.C3.B1a

Possible primary indicators

Digital Impact

Digital 'reach'

We can calculate with different approaches (depending on the features of the data served by associated data service) the popularity of different collaborative economy topics/services related to the collaborative economy, as well as its time-wise evolution

Some specific examples on how to arrive at this figure ('reach') in a digital channel:

1. 'Google Trends' results ('news' and 'search service')
2. Number of followers of the relevant account
3. Number of tweets per second containing any of the words in a 'bag-of-words' model
4. Wikipedia page-views of an specific article
5. Wikipedia edits of an specific article, although we think it's best to combine with the previous number (page-views) in a weighted way

It's important to note here that in order to 'normalize' these metrics, an absolute scale must be implement, and we must measure every occurrence against this scale. In each case, the scale reference levels must be placed analyzing well-known 'reference' services. For example, in the 'Google trends' metric, we must find two or three terms for reference and put the scale related to the values of that two or three terms (i.e: a word of maximum popularity, other of average popularity and other of very low popularity). In the case of Wikipedia, the scale must be implemented around two or three articles; and in the case of Twitter, of two or three accounts.

Existing communities on Twitter

'Listening' to different accounts and or hashtags linked with the collaborative economy we can build user networks (based on follow or retweet relationships) and work on the characterization of the network in order to know the different topologies implied. Arguably, the most useful network metric in this scenario is known as 'modularity' (how centralized or distributed is a network (https://en.wikipedia.org/wiki/Modularity_(networks)

Also, with this approach, we can measure community 'overlap', and see which populations and services intersect.

Another kind of analysis (more complex) could be what's called 'perspectivism': How many communities (and thus, perspectives) exist about a specific account or hashtag. An example of this approach can be seen at http://ceur-ws.org/Vol-1210/datawiz2014_07.pdf.

Digital controversy index

Studying discussion dynamics in Wikipedia, a 'controversy index' can be calculated for every relevant article (http://airwiki.ws.dei.polimi.it/index.php/When_the_Wikipedians_Talk), and thus, we can identify polemic articles using a range between 0 and 10.

On the other hand, listening to Twitter replies we can also measure satisfaction levels of users of a service, based on the number of complaints or compliments a brand or service is receiving in the form of replies.

Depending on the presence of both Wikipedia and Twitter for every term or service, we can use one metric or another (or both)

Market

Market index

¿What's the relationship between an asset cost (per unit or time) in the collaborative world vs the traditional market? This question cannot be answered directly by this specific metric (collaborative cost/traditional cost) in all sectors of the collaborative economy, but with some external stats, we can approach the number in different ways and scenarios:

- Uber 'trip' expenses vs a traditional taxi (normalized by kilometres) in the cities we are interested
- Same for Lyft trip expenses
- Airbnb expenses vs hotel expenses in the cities in which we are interested

Use

Approximate number of downloads for a specific 'client app' (if applicable)

In some services, we can know, with the help of stats like the ones provided by https://www.appannie.com/apps/google-play/app/com.airbnb.android/details/, the number of downloads of apps related to specific services, with additional time and geographic information. It's important to note that these numbers are an approximation, so you can only partially trust them

Absolute number of shared assets

If we have data about a specific 'service activity' (Airbnb 'listings', vehicle 'fleets' of 'Uber' and 'Lyft'), we can measure in an approximate way the rate of penetration (collaborative vs traditional assets) of a specific service in a geographic region, and track that throughout time.

Indicator combination

We propose four dimensions which will help in analyzing impact from each different service:

1. **Digital impact index:** Weighted average of four relative indicators: Google Trends, Twitter followers, Wikipedia edits and Wikipedia page-views
2. **Controversy Index:** Weighted average of Twitter and Wikipedia controversy indexes
3. **Market index:** "What's the relative difference between expenses in a collaborative economy service vs a traditional one?". That's the Market Index. It's very difficult to know the geographical segmentation in some cases in a direct way, but , it can be extracted in some, with lots of added value.
4. **Use index:** Weighted average of the number of app downloads and 'market penetration index (based on the ratio of collaborative assets vs traditional assets)

Global analysis

These four dimensions can be used to generate 'radar chart' or 'quadrants' (https://en.wikipedia.org/wiki/Radar_chart), useful as a starting point to characterize and compare different service and their evolution in time, enabling 'insights' such as:
1. In which moment in time are there spikes in the 'controversy index? Which kind of controversies are involved?
2. What kind of controversy has 'X' service compared with 'Y' service? What's the controversy level of 'X' service compared with 'Y' service? What are the differences?
3. What are the services with a better advantage against the market, from a user standpoint, with all the legal/ethical implications that this advantage can generate?
4. What are the services that are having a greater penetration and how will it look it the future?
5. Is there any relationship between 'Use Index' and 'Controversy Index'?

Geographical analysis

Indicators number 3 ('Market Index') and 4 ('Use Index'), in most cases can be geographically segmented, but Indicators number 1 ('Digital Impact') and 2 ('Controversy Index') lack this ability in most cases.They can be inferred in an approximate way (such as in Twitter and Wikipedia cases) , but not as solidly as in the case of the other two indicators)
Thus, we recommend to 'zoom in' the geographical analysis with only the help of Indicators 3 and 4 ('Market Index' and 'Use Index', respectively). These two Indicators can be segmented by service and cities, generating 'insights' such as:

1. In which cities the user is going to perceived a greater advantage using a specific service? (Indicator 3)
2. Is there any existing relationship between this aforementioned advantage and the rate of penetration of a service (Indicator 4)?
3. What's a possible classification for evolutions in time depicting 'Use Index' (#4) against 'Market Index' (#3) ?
4. What's the relationship between cost of life (from Numbeo or such) and the rate of penetration (Indicator 4)?
5. What's the 'disparity' level in 'penetration rate' (#4) in different cities of the same country/state/region?

Conclusions

Besides the scarcity of available resources (data) aimed at the characterization and modelling of collaborative economy services, we consider that in some cases (Uber, Lyft, Airbnb) there's enough data for applying it to analyze collaborative economy impact through the four proposed Indicators/Metrics

Otherwise, there's always an open road: asking for data to the services themselves (something similar was done here http://pentagrowth.com/), which (surprinsingly) led in some cases to new data revealed

Regarding indicators, we consider that these four dimensions of analysis measure four mostly independent variables of collaborative economy impact, and that with the help of local open data (specific for each particular study), new powerful insights can be generated, specially in the temporal or spatial dimension.

Chapter 3: Conclusions and further research

Goal

To present suggestions for further research on Sharing Economy's context, considering its use, available services and users. They're also shown some possible implementations for this research lines.

In the last part of this document we also present briefly some conclusions emerged from the current document.

Further research lines

Some action lines to extend the current investigation process are shown next:

- To monitorize national, regional and/or local open data repositories in order to detect new metrics or open datasets related with the Sharing Economy's activity. This can be reached by monitoring (automatically or not) the currently available open data services. This availability of metrics/datasets can be used too, as a partial metric of Sharing Economy ratio of use over a territory.

- To develop and manage data mining systems over the most relevant services (because of their activity volume, because of their activity context or just because they're great study cases) either through API access, if it's available, or using 'web scrapping' techniques for the data mining process. It can be done relaying over existing open platforms/software and/or developing own technical solutions in order to generate historical knowledge repository.

- To develop quantitative metrics from available data and to set references over time that can lead us to a better actors/services characterization and trends detection over either use ratios or market volume/offer. This can be reached by the obtention of some of the reference values commented over chapters 3, 4 and 5 and/or combining them in any of the proposed ways or in a new ad-hoc one.

- It would be also interesting, besides all the above, to carry an study in order to expose the Sharing Economy actors' coexistence dynamics. Do they establish better conditions for new sharing-actors and/or the rest of existing competitors? Do they tend to monopolize the market? This can be done using historical datasets comprehending various services in a particular geographic context.

Conclusions

Studying the Sharing Economy status and context is nowadays a complex task due to the next factors:

- Lack of transparency in the main (considering their economic strength) actors.
- Lack of data sources around the smallest actors. Most of the times it's hard even to find them.
- Lack of citizen/government awareness and/or legal tools on the local term in order to regulate the Sharing Economy market. There are still so few legislative initiatives to promote and to ease the public access to the data this Sharing Economy services generate.
- Lack of quantitative and/or official metrics/indicators in order to quantify services' penetration and impact. This scenario gets more remarkable over local contexts.
- Sharing Economy context is incredibly heterogeneous, and it's not always easy even reach consensus on what is or is not a collaborative service, or if it's a service belonging to 'Corporative' or 'Procommons' side of the Sharing Economy.
- Average live-periods of the services are pretty short (even when talking about segment leaders), which complicates their study and characterization.

Sharing Economy is a merging economic segment where new actors appears and old ones vanish in short time periods, where brand new uses over collaborative capital are accepted or disapproved constantly, and the legislation is still being thought in the vast majority of countries and cities or starts shyly to being implemented at some of them. It would be desirable a "cooling" process which could allow to build a basic typology of contexts or work areas on which the Sharing Economy services or strategies tend to be sustainable and/or generate profit.

It is likely to happen a boost on legislative processes around Sharing Economy worldwide, which will promote data coming from Sharing Economy services, platforms and entities to be published. That circumstance usually tend to encourage public analysis leading to firm conclusions about services' suitability and social return.

Some countries on the EU "space", where the Sharing Economy is reaching greater penetration ratios in classic economic markets and higher economic weight, start to study how to formally characterize and measure implantation rates, size of the SE services as well as political and social impact and consequences the SE could bring. (https://www.ons.gov.uk/economy/economicoutputandproductivity/output/articles/thefeasibilityof measuringthesharingeconomy/2016-04-05#toc)

Appendix I: Direct Data

Goal

List all data sources related to collaborative economy and describe them, including examples from both corporate collaborative economy and commons-oriented collaborative economy. Source: http://dimmons.net/wp-content/uploads/2017/01/Direct-Data.ods

Data format

Links to data sources have been retrieved, with additional information for each, structured in the following way:

- There are a couple of tabs in the spreadsheet: The first one ('corporate tab') with data sources related to corporate collaborative economy and the second one ('commons' tab) with commons data sources
- Field names for both tabs are as follows:
  - Service ('A' column): Service/entity name
  - Service type ('B' column): One of the following
    - Code sharing
    - P2P sharing
    - Ride sharing
    - Crowdfunding
    - Resource sharing
    - Goods
    - Vacation homes
    - Knowledge
    - Citizen labs
    - File Sharing
    - Shared media
    - P2P lending
    - Directory
    - Code sharing
    - Car sharing
    - Car rides

- Format: ('C' column): One of the following: CSV, JSON, XML, GeoJSON, TSV, SQL, Variable (in case of several or custom format), Visualization, Library (code) and Table (data can be retrieved with web scraping techniques)
- Type ('D' column): If data access is enabled through (a) Download, (b) API, (c) code o (d) HTML 'scraping'
- Size ('E' column): Approximate data size (if type is 'Download')
- Contents ('F' column): Brief description of the content of the data
- Url ('G' column): Data access URL
- Licence ('H' column): License type in case it's known (otherwise 'Unknown')

Appendix II: Participative data generation methods

Goal

To identify and locate experiences focused on generate listings/datasets/mappings of actions, events, services or resources around Sharing Economy environment. Likewise global resources regarding european space, either transnational experiences or source code repositories oriented to explore Sharing Economy services's data, will be listed.
Source:
http://dimmons.net/wp-content/uploads/2017/01/Participative-data-generation-methods-ENG.ods

A section focused on further explaining 3 prominent use cases is included:

- Camps carried on by Digital Methods Initiative
- 'MapJam' performed by Shareable people
- 'Labs' performed by MappingTheCommons

Format

Links to all the experiences, among additional information, have been gathered. The table containing the info is structured as follows:

- Promoter (Column 'A'): Entity promoting each listed experience. In data included on the table this field is given one of the following values:
  - Procomuns: http://procomuns.net/es/proposito/
  - Shareable: http://www.shareable.net/about
  - P2PFoundation: https://p2pfoundation.net/
  - MappingTheCommons: http://mappingthecommons.net/en/world/
  - ThePeopleWhoShare: http://www.thepeoplewhoshare.com/about/
  - DMI: https://wiki.digitalmethods.net/Dmi/DmiAbout
  - TransformMap: http://transformap.co/about/
  - Ripess: http://www.ripess.org/quien-somos/?lang=es
  - Individuals: If the author is a person or group of people which don't belong to an stable entity. It's used in two open source cases shown: Airbnb scrapping project

(Row 11 in the shared table) and Uber scrapping project (Row 12 in the shared table).

- Geographic Scope (Column 'B'): Geographic space on which the experience is carried on. In data included on the table this field is given one of the following values:
  - Barcelona
  - Amsterdam
  - Europa
  - Global

- Type (Column 'C'): Kind of experience depending on the event methodology. In data included on the table this field is given one of the following values:
  - Datathon: It's a collaborative data-generation-oriented activity around a particular theme. It usually develops on short time periods (24-48 hours). Methodologies vary but the core idea is to ease multidisciplinary groups in order to address the issue from different perspectives and approaches. The Datathon concept is based on Hackatons (https://en.wikipedia.org/wiki/Hackathon)
  - Collaborative Mapping: We tag experiences as "Collaborative Mapping" in the shared table if in spite of had been generated their results using community's work and knowledge, the won't fit on the Datathon concept previously described. Usually, the works is carried on individually and continuously in time.
  - Academic activity: These ones are experiences performed as university academic camps (Rows 7 and 8 in the shared table). Once again this format looks similar to the Datathon one, but not enough to include them on it.
  - Open Source Repository: Corresponding to rows 11 and 12 in the shared table. They are collaborative experiences which are performed over longer time periods, by an individual or small groups of people, not needing an explicit event or call for work. They release source which helps community in the data obtention, and/or analysis processes. This code can be modified, copied and mutated in order to generate own solutions.

- Url (Column 'D'): Experience's URL

Prominent use cases

Digital Methods Initiative Camps:

Twice a year Digital Methods Initiative performs an investigation Camp in order to ease and promote studies over digital sphere. For two weeks, participants generate dara, reports and open tools used on the process. Sharing Economy issue starts to get trendy on proposals, so some studies had yet been carried on. For example:

- In the 2015 Summer Camp a small toolkit for mining Github data as developed and released. It's still available at https://wiki.digitalmethods.net/Dmi/ToolDatabase . This set of tools was developed under a Data Driven Journalism project (https://wiki.digitalmethods.net/Dmi/GitHubDDJ). (some of these tools, which make an intensive use of resources need a formal petition for accessing)

- In the 2016 Summer Camp one of the teams carried on an study about Airbnb impact over Amsterdam city and its surrounding areas. (https://docs.google.com/document/d/1a4X0y1_1zsPMw5ieyAu71aHfNmfViFGg9NFt7ar13mc/edit#heading=h.5t7potm42uey) and they released a repository containing the source code used for data obtention from Airbnb systems. (https://github.com/DMI16-airbnb/dataTools).

MapJam

Shareable-net platform promotes the implementation of mapping events, regarding resources, activities and entities related with Sharing Economy. This events are called "MapJam". Shareable people offers expertise, documentation and communication support in order to carry on the event and map the Sharing Economy ecosystem over new locations. (https://docs.google.com/document/d/1hcxLZowxaXpE8FFcrflB8NEeB2XPKDwATdOR0k7QIn8/edit#)

They have managed to get commons resources mapped over 80 cities worldwide, most of them on Europe and North America. Depending on the city, different types and/or volumes of resources are shown. For example:

- París: https://docs.google.com/document/d/1fsdsqmBZ72WDGcROYvVvB56nj1Fvolq8kdI4Z2be63Y/edit#
- Barcelona: https://www.google.com/maps/d/u/0/viewer?mid=17lUXgv4a3Y19jXE1ny0CAL2CbHU

- Ljubljana:
  https://www.google.com/maps/d/u/0/viewer?mid=1_SppTkyUPehVy-mNiqwzhauAO5Q

MappingTheCommons

Transnational citizen project, focused on carry on resources mappings in the context of Sharing Economy. Their methodology implies to work on a selected city each time. (http://mappingthecommons.net/es/mundo/).

During the period 2010-2014 they had carried on various different events ('Labs') focused on cities across Europe and South America. Participants perform identification and mapping works over city's Sharing Economy resources. Visual and textual resources were also included on every 'Lab' in order to document and communicate the whole process. (for example: http://mappingthecommons.net/es/atenas/ o http://mappingthecommons.net/es/estambul/)

Appendix III: Possible statistical indicators

Goal

List all possible metrics/indicators related to collaborative economy, derived from statistical sources (e.g: Eurostat)
Source: http://dimmons.net/wp-content/uploads/2017/01/Possible-statistical-indicators.ods

Format

Two types of data have been retrieved: (a) Data sources and (b) Possible/Relevant (i.e: specific) data

In the first case, the goal has been to bring coverage on different levels: International, European, and local (specific cities). As for the second case, the intention has been to choose several data 'categories' that can provide context that help studying collaborative economy impact

In the corresponding spreadsheet there are two tabs:

1. 'Sources' tab, with the following columns:
   a. 'Name': name of the specific source/institution
   b. 'Source': URL that points to the resource
   c. 'Scope': 'All' or 'Development' (data relate to the status of development of countries)
   d. 'Geographical scope': Specific world area covered by the source
   e. 'Description': Short description of data contents
2. 'Indicators':
   a. 'Name': of the indicator (or of the dataset used to extract indicators)
   b. 'Source': URL that points to the resource
   c. 'Scope': Conceptual scope of the data. One of: Economic Development, Education, Equality, Demographics, Digital Development, Economics, Governance, Research, Environment, Globalization, Transport, Life Quality, Culture, Tourism, Innovation, Lodging, Associations or Social Services
   d. 'Geographical scope': Specific world are covered by the source; All, specific countries, OECD countries, Europe, Catalunya, Barcelona
   e. 'Description': Short description of data contents