

# Data science for perception

Design of data driven pipelines in the matters of human  
perception



**Sagar Joglekar**

Department of Informatics, Faculty of Natural and Mathematical Sciences  
Kings College London

This dissertation is submitted for the degree of  
*Doctor of Philosophy*

March 2019



*To Geetika, Medha and Chanda ... The three formidable pillars of my life*



## DECLARATION

---

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other University. This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration, except where specifically indicated in the text. The dissertation covers contributions for journals and conferences where I was the main contributor and primary investigator.

Sagar Joglekar  
March 2019



## ACKNOWLEDGEMENTS

---

And I would like to acknowledge ...



## ABSTRACT

---

The idea of wisdom of the crowds has been tested time and again when it comes to applications like recommendations of items or places, which inherently taps into the likes and dislikes of the crowds. Interesting convergence has been shown to emerge when insights are driven by usage patterns of large number of people.

With the ever pervasive nature of the internet, we as a society have started treating the online spaces not only as a tool to access information seamlessly with atomic transactions, but also as a natural extension of the self. We spend more time than before as a part of a larger networked community, exchanging thoughts, debating ideas, expressing creativity, empathy and sometimes seeking help. The presence of these extended set of interactions over the web, implies that an intangible yet essential affective component is now driving our behaviour when we empathise with people online, or interact with creative or aesthetically pleasing content. At such a juncture, I ask the question "Can we build pipelines that look at these interaction patterns, to understand and utilize the perception driven behaviours of humans on the web?"

In this dissertation I explore the idea of "Perceptions of the crowds" through two case studies. I build systematic pipelines that quantify human perceptions in both cases interpreting interdisciplinary ontologies. In the first study I analyse on-line formal networks where interactions between humans are purely with the aim of helping each other. I develop frameworks to abstract out the structure of these conversations in computable structures. I then delve into investigating presence of supportive phenomenon by finding discriminative local structures in these abstractions. I reason about these local structures using established cross disciplinary theories. This informs my analysis about the nature of peer to peer support in these communities and paves the way to do actionable intervention in the area of peer support in online networks.

In the second study, I investigate utility of perceptions of physical spaces, by developing a pipeline that capitalizes on perception of urban aesthetics at the crowd scale, to develop intervenable insights around urban beautification. I propose a deep-learning driven framework, which is able to quantify the perception of intangible abstract qualities like 'beauty' through

a crowd sourced rating of google street view images. I show that a general pattern of beauty in urban spaces can be learnt through a crowd sourced opinion and deep learning models. I further develop a generative model to simulate beautification of urban spaces. Through a detailed literature review of the field of urban design, I develop a measurement framework which can provide insights into the predictors of urban beauty on a case by case basis using well known urban design metrics. I validate the value of these metrics through expert survey and validate the interventions through crowd sourced perception experiment.

As an over arching goal, through this dissertation, I try to build a repeatable framework for data driven pipelines that can capitalize on data that contains human perception signals. I propose that building custom representations that bind computational metrics with interdisciplinary ontologies is a necessity to build useful pipelines.



## TABLE OF CONTENTS

---

<b>List of figures</b>	<b>xiii</b>
<b>List of tables</b>	<b>xv</b>
<b>Nomenclature</b>	<b>xv</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Introduction . . . . .	1
1.2 Perception and Affect . . . . .	2
1.3 Intervention and the DIKW model . . . . .	3
1.3.1 Data . . . . .	4
1.3.2 Abstractions . . . . .	5
1.3.3 Knowledge . . . . .	5
1.3.4 Wisdom . . . . .	6
1.4 Research hypothesis . . . . .	6
1.5 Thesis overview . . . . .	7
1.6 Original contributions . . . . .	8
1.7 Outlook . . . . .	8
1.7.1 Notes . . . . .	8
<b>2 Online health forums primer</b>	<b>9</b>
2.1 Dataset and properties . . . . .	11
2.2 Graphs: A primer . . . . .	11
2.3 Activity patterns of users . . . . .	11
2.4 Dataset characterzation . . . . .	13
2.4.1 Activity . . . . .	14
2.5 Propensity to help . . . . .	15
2.6 Not like conventional networks: Anti-rich club effect . . . . .	15
2.7 Key takeaways, possible interventions . . . . .	15

<b>3 Structure of online supportive conversations</b>	<b>17</b>
3.1 Introduction . . . . .	17
3.2 Results . . . . .	19
3.2.1 Peculiarity of threads of Support . . . . .	20
3.2.2 Patterns in local interactions . . . . .	22
3.3 Methods . . . . .	23
3.3.1 Data . . . . .	23
3.3.2 Abstractions . . . . .	24
3.3.3 Macro and local metrics . . . . .	26
3.3.4 Structural metrics . . . . .	27
3.4 Appendix . . . . .	28
3.4.1 Triadic statistics for twitter conversations . . . . .	28
3.4.2 . . . . .	28
3.4.3 Network characteristics . . . . .	28
<b>4 Perceptions in real spaces</b>	<b>33</b>
4.1 Introduction . . . . .	34
4.2 Related Work . . . . .	35
4.3 FaceLift Framework . . . . .	36
4.4 Evaluation . . . . .	42
4.5 Conclusion . . . . .	50
<b>5 Perception and generative models</b>	<b>51</b>
<b>References</b>	<b>53</b>

## LIST OF FIGURES

---

1.1	The DIKW pyramid . . . . .	3
3.1	Panel shows CDFs of different network metrics. Fig.3.1a shows the response time distributions, Fig.3.1b shows symmetrically engaged users, Fig.3.1d shows topical similarities across posts and 3.1e shows the branching factors of reply graphs. . . . .	21
3.2	This panel shows the statistical significance of the three over expressed and one under expressed triadic motif. . . . .	23
3.3	Example UserGraphs and their corresponding Reply graphs, Figure 3.3b shows a random thread from the SW sub-reddit and 3.3a shows the corresponding reply graph that arises from the response structure of the same thread. In comparison we have Usergraph Fig 3.3d and its corresponding reply graph Fig 3.3c from one of the Front page threads . . . . .	25
3.4	Figure 3.4a shows the 16 different types of motifs that are looked for in the user graph data. Figure 3.4b shows how three unique users could produce different motifs. The three shapes represent different users and the dotted line means the message order is irrelevant. . . . .	29
3.5	The figure shows comparison of occurrence ratios of 9 insignificant motifs. Blue traces are for Suicide watch and Green traces are for Baseline Front page threads . . . . .	30
3.6	<i>Fig 3.6a shows the branching factor for twitter threads that talk about suicidal tendencies against baseline threads. Fig 3.6b shows the distribution of median centralities per thread, for both the twitter crawls. Fig 3.6c shows Distribution of symmetric messages in reply graphs for both datasets. Fig 3.6d shows the distributions for users participating in a symmetric conversation Fig 3.6e shows the distribution of reply urgency for suicide threads against baseline. The suicide median reponse time for suicide threads is 3 min as compared to 18 mins for non-suicide threads . . . . .</i>	31

3.7	<i>Fig 3.7a shows the distribution of maximum depths of Reply Graphs for Subreddit r/SuicideWatch and the baseline Frontpage conversations. Fig 3.7b shows the distribution of unique authors per thread in the two datasets. Fig 3.7d shows Distribution of degrees for Reply Graphs, r/SuicideWatch and FrontPage. Fig 3.7c shows the degree distributions for the reply graphs . . . . .</i>	32
3.8	<i>Distribution of responses per thread on Subreddits r/SuicideWatch and Frontpage . . . . .</i>	32
4.1	A simplistic end to end illustration of the FaceLift framework. . . . .	37
4.2	Frequency distribution of beauty scores. The red and green lines represent the thresholds below and above which images are considered ugly and beautiful. Conservatively, images in between are discarded. . . . .	39
4.3	Two types of augmentation: (a) rotation of the Street Views camera (based on rotation); and (b) exploration of scenes at increasing distances (based on translation). . . . .	39
4.4	The types of scene that have greater propensity to be correctly augmented with similar scenes at increasing distances. . . . .	40
4.5	Number of labels in specific urban design categories (on the $x$ -axis) found in beautified scenes as opposed to those found in uglified scenes. . . . .	45
4.6	Count of specific walkability-related labels (on the $x$ -axis) found in beautified scenes minus the count of the same labels found in uglified scenes. . . . .	46
4.7	The percentage of scenes (y-axis): (a) having an increasing presence of sky (on the $x$ -axis); and (b) having an increasing level of visual richness (on the $x$ -axis). The error bars represent standard errors obtained by random re-sampling of the data for 500 iterations. . . . .	47
4.8	Interactive map of FaceLifted scenes in Boston. . . . .	49

## LIST OF TABLES

---

3.1	Notations and Terms. . . . .	19
4.1	Notations . . . . .	36
4.2	Percentage accuracy for our beauty classifier trained on differently augmented sets of urban scenes. . . . .	38
4.3	The table showcases examples of the “FaceLifting” process. It is worth observing that the process of beautification prefers greenery, narrow roads and pavements . . . . .	43
4.4	Urban Design Metrics . . . . .	44
4.5	Coefficients of logistic regressions run on one pair of predictors at the time. . . . .	48
4.6	Urban experts polled about the extent to which an interactive map of “FaceLifted” scenes promotes: (a) decision making; (b) citizen participation in urban planning; and (c) promotion of green cities . . . . .	49



# CHAPTER 1

## INTRODUCTION

---

*One should never try to prove anything that is not almost obvious - Alexander Grothendieck*

### 1.1 Introduction

We live in a world where information is being bombarded on our cognitive faculties from all sides, at all times. The internet is a continuous stream of information and each source is fighting with the other to get a piece of our attention budget. With the advent of machine learning and big-data sources, building systems that predict actions as a response to perceptual triggers is the bread and butter of many companies. The use cases may range from understanding which advertise made a visitor do an unscheduled purchases on amazon, or which string of music tracks recommendations maximized a users time on a particular music platform. But in the end it all boils down to understanding what triggers result in human action or lack thereof[44]. Nonetheless the systems that surrounds a human interacting with the internet, are all figuring out the best triggers which are perceived by the human as worthy of attention. The term “Attention Economy”[8] was actually coined for this very reason. In the words of Matthew Crawford “*Attention is a resource, a person has only so much of it.*”[7]. We live in the age of distraction, and more often than not, our perceptions are guiding our actions, than our cognitive processes. Several studies have shown that engagement is almost always a game of stimulating our most basic urges, such as dopamine hits, presence of faces or simply arousal of emotions to increase the working memory.[3, 21][40][43]. An interesting side effect of dwindling attention budgets is the emergence of more formal topical spaces on the internet. The ever pervasive nature of the internet allow these formal spaces to function almost like physical communities, with moderated and effective peer to peer exchange of thoughts, ideas and empathy[19, 25, 46].

In such an environment, as computer scientists, it is worth asking the question: **Can we develop pipelines that can look at the structures in the data to quantify how humans perceive their sphere of interaction? Can we use that perception to design impactful interventions towards well being of the users, on and off the internet?**. This question can be decoupled and decomposed into two fundamental problems:

1. How do human perceptions manifest in data?
2. How do we quantify these perceptions?

These two questions are going to be the guiding principles of my dissertation.

But first of all, we need to clarify the relation between perception, affects and intervention and to do so we should try and understand each of these terms separately as a part of the native field's context.

## 1.2 Perception and Affect

This paragraph needs to be refined with more literature that convinces the reader that perception -> emotion. And any knowledge extraction pipeline inherently links the two together. In this dissertation, I would try to build frameworks to capture human perceptions in the realm of human to human interactions and subjective aesthetics. The utility of such an attempt, can only be justified if there is a real link between how humans function at the most fundamental cognitive level and how they perceive the intangible, including the aesthetic. There has been an ongoing effort to unravel this link, through psychological, neuro-evolutional and philosophical arguments. I will try to gain inspiration from them, but a detailed critique is beyond the scope of my dissertation and expertise.

**Affect (APA definition):** *any experience of feeling or emotion, ranging from suffering to elation, from the simplest to the most complex sensations of feeling, and from the most normal to the most pathological emotional reactions.*

**Perception (APA definition):** *the process or result of becoming aware of objects, relationships, and events by means of the senses, which includes such activities as recognizing, observing, and discriminating. These activities enable organisms to organize and interpret the stimuli received into meaningful knowledge and to act in a coordinated manner.*

Emotions or ‘affects’ and perceptions have long been discussed in the psychology, neuroscience and philosophical literature. Emanuel Kant in his prolific work, first discussed the utility and the philosophical reasoning behind presence of affects or emotions[22]. In his opinion, emotions are pre-cognitive involuntary states, termed as "mere perceptions of

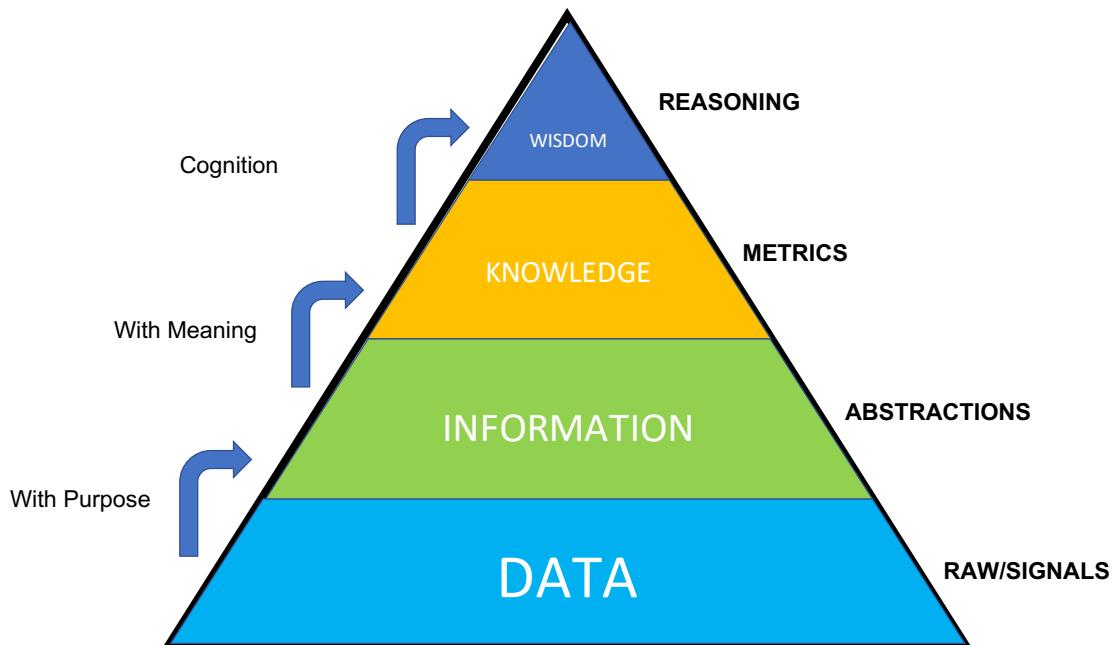


Fig. 1.1 The DIKW pyramid

unspecified bodily states"[5]. But that does not mean they don't influence our deepest level of well-being. The link between affect and perception has also been explored in several cases. An argument to link perception of affect arousing aesthetics was made by Perlovsky[34], where they propose that the phenomenon of affects arousing from aesthetics, comes from a fundamental human need enrich the knowledge bases about real world. An unexpected thing of structure, satisfies this need at some level and we perceive it as aesthetically pleasing. Another recent study by Zadra et.al[50] evaluated the relation between visual perception and emotions. They demonstrate that the conventional assumption of the disentangled functioning of perception and affects is not necessarily true. Humans are quite susceptible to perceiving different realities based on different aroused affects.

So the community is still unclear, but there seems to be a theory that

### 1.3 Intervention and the DIKW model

The reflection to find the answers the questions, ties us back to one of the fundamental frameworks about operations on data, that is the **Data , Information, Knowledge and Wisdom** model[37].In this dissertation I posit and demonstrate through case studies, that for reasoning about perceptions, you need to interpret the metrics that contain contextual knowledge about the abstractions you are working with through the lens of the target ontology. Debating the nature of this ontology is not the purpose of this dissertation, but I show that 'a'

solution can be reached, if systems are designed to interpret metrics by anchoring around an ontological bias.

In this model, the most base layer consists of the pure form raw **data or signals** that come from a source. If we are measuring perceptions of human, this source needs to be tied back to humans in some way. The data needs to be generated as a result of some human-human interaction or as a result of deliberate human input as a form of response to their perception of some event.

The **information** layer is the result of the fact that any process done on the data is with a sense of purpose or an end goal. For example, if the goal is to understand how humans exchange messages at times of distress, you would most certainly need to express the raw information about sender and recipient of messages into some form of a networked abstraction. The abstraction preserves the organization of data, but at the same time allows information to be operated on. As a result, almost always the output of this process is some form of a data abstraction. In the next stage of processing, you need to attach some meaning to the patterns in the information to extract

**knowledge**. This For example, if you need to know the most popular user who has exchanged the most messages with others among a network of users exchanging messages; you would look for the most central user in the network. In this particular case, the metric of centrality dawns the meaning of popularism in the context of our message network.

The final layer needs a cognitive process and an ontological framework, to extract actionable insights, which we can call **wisdom**, from the knowledge. By classical definition of ontology, it defines properties of and relations between objects. For this very reason, these ontological frameworks need to be originating from the field of intervention. In our example, a human analysing data needs to get some insights about the dynamics of popular users which requires the human to consume the metrics and derive insights. However the insights need to be grounded in psychological ontological framework, for us to derive certain trends in behavioural dynamics of this network. Figure 1.1 shows an illustration of the adopted version of Rowley's DIKM model.

### 1.3.1 Data

As discussed, this dissertation is about developing frameworks around quantification of human perceptions, such that we can do impactful interventions from this approach. And the most base level of this pyramid is the data that the frameworks would work with in order to progress on these lines. I work with two forms of data, textual data and image data, to understand two separate forms of perceptions.

### **Textual Data**

The first case study of this dissertation focusses on online support communities. It has been shown through several studies in medical informatics, that these communities play a very important role in providing support and respite in times of distress. The communities are especially helpful when it comes to people suffering from long term illnesses or mental health issues. To understand how users on these communities perceive social support, I work with data acquired from online health forums, where users share, give support and ask for support. I look at communities that deal with long term conditions like Lung illnesses, and communities where mental health patients seek support.

### **Image data**

The other facet of my work looks for quantification of how we perceive physical spaces. I work with google street view data, and the aim of this work is to understand, through crowd sourced methods, how people perceive the sense of beauty in urban areas.

### **1.3.2 Abstractions**

As mentioned before, the act of aggregating information from data, almost always involves building organized abstractions. In case of the first study with textual data, I incorporate user information into the textual data to build organized networked structures, which can then be evaluated using graph theoretic methods. To understand textual patterns, I use the language embedding models which would be discussed in the chapters to come. While processing image data, I use several segmentation techniques to group semantically similar pixels together. I also use several state of the art object and scene detection to extract semantic information from an image. A more detailed discussion of these abstractions would be done in the later chapters.

### **1.3.3 Knowledge**

For extracting knowledge, we need to first associate meanings to certain computable metrics that we obtain from the abstractions. As discussed in the previous example, it could be as simple as associating the property of “popularity” to the metric of centrality. In my case, I develop several of these metrics for the two studies, some based on intuitions which I test validity for, and some based on extensive literature survey. To give an example, I develop the concept of anchored triads, which combines local interaction motifs with the role of a user in a supportive conversation to understand how these conversations evolve.

### 1.3.4 Wisdom

Finally the wisdom in my case, is simply the philosophical, quantitative and qualitative discussion about what perception did we actually capture from these progressive operations. To answer the questions posed by this layer, we need to hold on to **an ontological framework** which grounds the metrics in the field of human interaction and planned intervention. This ontological framework that deems meaning to the structure in data, needs to be achieved through cross disciplinary literature review. In case of social support, what metrics combined with the understanding of human behavioural ontology, give rise to signatures of social support? Could those be found in sociological literature that distils social ties as interaction motifs? Or can that be found in pure statistical understanding of networks and formation of edges? And in the end, can we provide valuable interventions in the scenarios where humans are asking for help online. My dissertation navigates similar questions to arrive at 'an approach' to quantify and intervene using signatures of human perceptions from data.

## 1.4 Research hypothesis

The global hypothesis of my dissertation as described before comprises of asking the two aforementioned questions: **How do human perceptions manifest in data?** and **How do we quantify these perceptions?**. But these questions are quite open ended, and answering them in a generalized manner seems impractical. For this reason, I need to first contextualize my work in the realm of practical applications. To rationalize the pursuit of these questions, I choose application driven case studies, which allow me the luxury to do a data driven exploration with a final goal of designing intervenable frameworks. Across my investigations, I follow the DIKW framework layer by layer, by distilling actionable wisdom from data. Through the two case studies , my work touches a diverse set of data science tools which range from complex networks formulations to generative adversarial models of human perception.

My research explores the following global hypotheses using the two case studies.

**H1: Data about human interactions from the web can be used to capture signatures of perceptual processes**

**H2: Human interactions data from the web can be used to quantify subjective attributes of the real world**

**H3: These quantifications can be effectively used to drive intervenable insights to improve perceptions of the said humans.**

Human interactions data from the web can be used to capture peculiar formats of human conversations

Ontology in a philosophical context is a branch of metaphysics, which asks the questions like "What exists?", "What comprises an object? ", "What hierarchies of object classes are present?", "What different entities form a higher entity?". Naturally this framework was suitable in information sciences to describe datum as hierarchies or taxonomies of other objects.

## 1.5 Thesis overview

In the first study, I examine the structure behind how people seek engagement on-line. I discover that humans are very limited by their attention budgets, and engage with informal social networks in a very atomic and primacy driven way, which makes prediction of engagement using certain attributes of the content very feasible. But on the contrary, in more formal social networks, where the aim is to exchange knowledge and support, the network evolves around certain key elements of perception of support and helpfulness, which are well explored in the real world communities. On the journey to unravel these traits, I develop techniques to abstract out online formal conversations and develop a models for detecting supportive conversations on the web. I discuss the utility of such a model and draw parallels with the offline model of community support.

In the second study, I investigate utility of perceptions of real world places through a crowd sourced rating of google street view images. I develop models to extract the perception of the crowds using data driven inference methods. I show that a general pattern of beauty in urban spaces can be learnt through a crowd sourced opinion and based on this finding, I develop a generative model to simulate beautification of urban spaces by using deep learning. I validate the quantification of perception of real-world beauty using crowd validation. I contribute a way to use computer vision techniques to abstract out beautification process into explainable metrics used by architects and urban planners. The final contribution is a demo

web application, that allows practitioners to examine and validate the utility of such a end to end system that captures citizen perceptions for urban design.

## 1.6 Original contributions

## 1.7 Outlook

### 1.7.1 Notes

Some points to cover in introductions

- Kant's theory of emotions , and how stimulus -> cognition -> emotion and action links everything together
- importance of capturing affects (desire to influence behaviour)
- There have been attempts to capture this , and how the plutchik's frameworks has helped in distilling certain emotions and capturing them from different forms of media
- this dissertation takes an different approach, in which we try to find signatures of affective responses seen online
- the dissertaion works in two parts, the first part looks at how support can be detected from structures in coversation
- part 2 looks at how the affect of beauty impacts how citizens perceive cities and can we capture that affect to improve urban design process

# CHAPTER 2

## ONLINE HEALTH FORUMS PRIMER

---

Online communities have the potential to influence health and health care. Recent studies have suggested that the participation of people with long-term conditions (LTCs) in online communities (1) improves illness self-management [1], (2) produces positive health-related outcomes [2-4], (3) facilitates shared decision-making with health care professionals [5,6], and (4) may even reduce mortality [7].

There is also evidence that self-management support interventions can reduce health service utilization [8,9].

Online communities have experienced an upsurge in popularity among people with chronic respiratory conditions such as cystic fibrosis [10], asthma [11], pulmonary hypertension [12] and chronic obstructive pulmonary disease (COPD) [13]. More than 15 million people in England suffer from a long-term condition or disability, and they account for at least 50 percent of all general practitioner appointments [14,15]. Thus, assessing how these online communities function and evolve can have important implications for health care provision.

This form of “user-led self-management” of LTCs bears similarities with the “expert patient” model, an approach to self-management of LTCs produced by the United Kingdom (UK) Department of Health in 2001 [16]. Evidence of the effectiveness of conventional off-line self-management programs based on the expert patient model, though, has been weak [17]. Clinic-based self-management programs often failed because of: (1) lack of awareness and engagement among patients and staff, (2) failure to consider low health literacy or cultural norms, (3) lack of attention to the need for family and social support, and (4) a fragmented approach to the provision of health and social care [18]. Although online health communities can be seen as an extension of the expert patient model, network effects, in addition to the online disinhibition effect [19], make them a distinct and unique complex intervention mechanism.

On average, one in four people with an LTC who use the Internet tries to engage online with others with similar health-related concerns [20]. In particular, it has been suggested that the value of participating in an online community lies in the possibility of gaining access to a range of people and resources quickly, easily [21], and anonymously [4], as well as obtaining tailored information and emotional support [1,22-26]. However, most of this evidence comes from qualitative studies [1,27], whereas only recent years have witnessed an increasing interest in quantitative assessments of online communities as intervention mechanisms [28-33]. Recent studies have been concerned with the users' unequal contributions and engagement patterns, and with the role of superusers. However, the contribution of superusers to the sustainability of online health communities and their structural properties remains mostly unclear.

The potential future integration of online health support systems with formal health care provision should be underpinned by a better understanding of how they are used and by evidence of their effectiveness. Indeed, as suggested by the Medical Research Council [34], integrating online support systems with the more traditional health care provision would require the identification and comparative assessment of potential alternative intervention mechanisms.

An expanding body of literature concerned with social network analysis has examined the structural patterns of relations among interacting actors and the social mechanisms that enable them to gain access to valuable resources [35]. There is also increasing evidence that network approaches can be applied to understanding the users' "expertise" [36], their interactions, and network effects on health-related outcomes in online health communities [37,38]. Uncovering the mechanisms underlying the formation of successful social networks requires a study of how online connections among people, namely the social ties or links, emerge and evolve, and how groups of individuals gradually grow in membership and become interconnected with one another. These processes of tie creation and group formation in online patients' communities are still mostly unexplored [1].

In this study, we performed a network analysis of the structure and dynamics of two online communities of people with LTCs. We chose the Asthma UK and the British Lung Foundation (BLF) communities as an exemplar of such communities because their users typically suffer from chronic respiratory conditions. In particular, while Asthma UK users typically suffer from a respiratory condition characterized by variable and recurring symptoms, BLF users represent a more heterogeneous population of participants affected by different diseases linked to chronic symptoms of breathlessness (eg, COPD, pulmonary fibrosis, cystic fibrosis, and lung cancer). We aimed to uncover and understand how these communities function and evolve, and the role that some users have in maintaining integration and cohesion (see

Textbox 1 for research questions). Ultimately, this study provides evidence for gauging the effectiveness of different interaction patterns and the users' structural positions and their potential for enhancing and sustaining health online communities as scalable self-management support interventions.

## 2.1 Dataset and properties

Data were collected by HealthUnlocked [39], the online platform provider of the Asthma UK and BLF communities. Registered users can choose to either write posts publicly or send private posts to one another. In the latter case, posts are shared between 2 users only, whereas when posts are written publicly, a large number of users can become connected through threads of posts. Only posts that were shared publicly were collected and analyzed. For this study, user identifiers (IDs) were anonymized by HealthUnlocked, and no demographic information was collected. The data sets included posts and their metadata (ie, the anonymized user ID numbers), user roles (eg, user, administrator, or moderator), date of posting, the hierarchical level of the post within the corresponding thread, and the dates in which the users joined and left the community. Both communities were moderated, and HealthUnlocked moderators (identified through metadata linked to posts) were included in the analysis to assess their contribution and compare it with other users. Online communities on the HealthUnlocked platform benefit from additional functionalities compared to other online forums, such as built-in patient groups that moderate the content. In particular, the content accessed by users is tailored to their interests, and profiles highlight users' condition, chosen community, medications and treatments they use or find interesting. No data were collected on participants' characteristics, though only people declaring themselves to be older than 16 years were permitted to create an account and take part in the online communities.

## 2.2 Graphs: A primer

### 2.3 Activity patterns of users

We looked at the number of users, the number of posts and connections per user and posting frequency. A connection (ie, a tie, link, or edge) was established from one user to another when the former replied to a post by the latter (see Textbox 2 for network analysis terminology). The pattern of connections generated over time through the cumulative number of posts and replies was examined. We were interested not just in the number of posts and responses but in who responded to whom, and when. To this end, we used social network

analysis [40] to visualize and study the structure of the relationships between users. Both visualization and analysis were conducted using the Gephi software. The network analysis was carried out through additional custom computer code in python. Descriptive analysis of the networks (ie, number of users, posts, and posting frequency) were calculated using the Pandas library, an open source library providing data structures and analysis tools for the Python programming language.

As a result of the small percentage of users who wrote posts to a disproportionately high number of users, the users' activity showed long-tailed distributions. Therefore, our analysis was based not only on means and standard deviations but also on medians.

To uncover time patterns in posting activity, we used Fourier transforms of the time series of the users' activity [46], a known method used for the analysis of signals. Through Fourier transforms, we identified the frequency components, called harmonics, that together made up the posting activity stream. In other words, we regarded the posting activity over the entire observation period in both communities as a complex signal and identified the frequency components that made up such a signal. This analysis was performed using custom code in Scipy, a Python-based scientific computing library.

The “rich-club” coefficient is a metric designed to measure the extent to which well-connected users tend to connect with one another to a higher degree than expected by chance [43]. To this end, for each value  $k$  of a node's degree (ie, the number of other users a given user is connected with), we computed the ratio between the number of actual connections between nodes with degree  $k$  or larger and the total possible number of such connections [47]. We then divided this ratio by the one obtained on a corresponding random network with the same number of nodes and degree distribution (ie, the probability distribution of the degrees over the whole network) as the real network, but in which links were randomly reshuffled between nodes. Thus, the rich-club coefficients may take values lower or higher than 1, depending on whether the real network has a higher or lower tendency to coalesce into rich clubs than randomly expected. In particular, networks that display a high rich-club coefficient (ie, greater than 1) are also said to show a “rich-club effect,” namely the tendency to organise into a hierarchical structure in which highly connected nodes preferentially create tightly knit groups with one another, thus generating exclusive clubs of (topologically) rich nodes, as illustrated in previous work [48].

In our study, superusers were defined according to their cumulative activity over the entire observation period. In total, we identified 400 superusers. To uncover how many superusers were active within each week, we detected how many unique users, among the 400 identified over the entire period, were active within that time window.

Following Zhang et al [36], the “z-score” was used as a proxy for users’ expertise. According to this measure, replying to many questions suggests one’s expertise, while asking questions indicates lack of expertise. In our analysis, we treated anyone starting a thread as a help-seeker, and anyone commenting on the thread as a help-giver [36]. Accordingly, the proposed z-score aims to capture the combined help-seeking and help-giving patterns. To this end, for each user, we measured how many standard deviations the observed total number of the user’s help-giving posts lies above or below the expected number of help-giving posts for the whole system. We extended the approach proposed by Zhang et al by empirically assessing the probability of posting and answering a question across all users over the entire observation period. In the BLF community, we found that the probability of answering is  $Pa=2/3$ , while the probability of posting is  $Pq=1/3$ . We assumed a Bernoulli process of posting an answer or a question to the forum, with probabilities defined as above. The z-score for a given user  $i$  was calculated according to equation (a) in Figure 1, where  $ai$  refers to the total number of answers user  $i$  posted to the forum,  $qi$  is the total number of questions user  $i$  asked in the forum, and  $ni=ai+qi$  is the total number of messages posted by user  $i$ . To obtain  $Zscore_i$ , let us define a random user that posts the same total number of messages  $nrandom$  to the forum as user  $i$  (ie,  $nrandom=ni$ ). We would expect this random user to post an average number of answers to the forum given by equation (b). Plugging in the value of  $Pa = 2/3$ , we obtained equation (c). Similarly, we would expect the random user to post answers with a standard deviation given by equation (d). Plugging in the value of  $Pa = 2/3$ , we obtained equation (e). To measure how many standard deviations above or below the expected random value a user  $i$  lies, we then computed  $Zscore_i$  according to equation (f). Plugging in the values of  $\mu_{random}$  and  $\sigma_{random}$ , we obtained equation (g). Finally, by substituting  $ni = ai + qi$ , we obtained equation (h).

## 2.4 Dataset characterarization

The data sets span, respectively, 10 years for the Asthma UK and 4 years for the BLF communities (see Table 1).

Despite the shorter time span, as a result of the larger number of users, the number of posts in the BLF community was higher than in Asthma UK, namely 875,151 compared to 32,780 respectively. Moreover, BLF users wrote a higher number of posts per user and were connected with a higher number of other users when compared with people in the Asthma UK forum (see Figure 2). In both communities, 60%-70% of registered users wrote no posts (ie, they were lurkers). Users who wrote more than one post contributed with a median

of 8 (range 2-8947) and 5 (range 2-1068) posts in the BLF and Asthma UK communities, respectively.

The number of official moderators among the highly active users was negligible; there were no moderators in the top 5% contributors to BLF and only 2 in the top 5% for Asthma UK. Thus, our network analysis predominantly reflects content originated from registered users.

When classified according to posting activity (ie, number of posts written to the forum), the top 5% users contributed to a substantial proportion of all posts: 58% and 79% in the Asthma UK and BLF communities, respectively. Superusers were those who made a high number of connections with other users in both Asthma UK and BLF communities (see nodes of large size in Figure 2). Asthma UK superusers made a lower number of connections than BLF ones. The posting activity of these superusers will be analyzed in more detail in subsequent sections.

## 2.4.1 Activity

### Posting Activity

The cumulative number of messages posted grew uniformly over time in the BLF community. By contrast, in 2015, the Asthma UK forum witnessed a substantial increase in posting activity, at a time coinciding with its move to the HealthUnlocked platform (see Figure 3A and B). This increase in activity can be attributed to the online community functionalities offered by HealthUnlocked, as described in the Methods.

The number of posts per user per week oscillated around a decreasing and an increasing trend (Figure 2C and D), while at the same time the number of posts always went up over the study period (Figure 1A and B). This suggests that there were intervals of time during which the rate of increase in new users was larger than the rate of increase in total posts. Moreover, in the Asthma UK forum users wrote according to two time patterns—they posted at an interval of 1-20 days or 6 months (Figure 4A), while those in the BLF community at an interval of 2 days (Figure 4B).

As more users joined the communities and connected to one another through online posts, distinct groups of connected users started to emerge. These groups, called network components (see Textbox 2), have fundamental implications for the effectiveness of processes of network dynamics such as information diffusion [49]. In a relatively short period, both communities underwent the formation of the “largest component” of connected users, namely a connected subset of users whose size increasingly outgrew the size of all other components (see Figures 1 and 4, and Multimedia Appendices 1 and 2). The largest connected components in both communities included 60

Figure 5 suggests that, as time went by, the number of forum participants and their posting activity increased, and the proportion of users who were part of the largest components decreased. This finding was expected because the number of posts also rose exponentially, yet at times at a lower rate than the one at which new users joined the communities (see Figure 1C and D). It, therefore, became more difficult for the network to self-organize into a connected component that would include 100% of the users. Figure 5A also shows that around week 450, when the forum moved to the HealthUnlocked platform, a larger fraction of users began to join the largest connected component, thus highlighting the role that the new online platform played in strengthening the connectedness of the network (see also Figure 3A and B).

## **2.5 Propensity to help**

## **2.6 Not like conventional networks: Anti-rich club effect**

## **2.7 Key takeaways, possible interventions**



# CHAPTER 3

## STRUCTURE OF ONLINE SUPPORTIVE CONVERSATIONS

---

### 3.1 Introduction

The world has become more connected over the past decades thanks to the networked nature of the technologies of the day. It is seldom possible to spend a whole day without single interaction on the Internet. The Internet gives platforms where we can not only connect with our social counterparts but also exchange ideas and express opinions. These new mediums have become so ubiquitous, that some research suggests that they might be affecting our broader psychological state [6]. But on the positive side, studies have also proposed different ways in which this medium could be used for measuring and intervening in the matters of mental health[10, 11]. Online communities, or *fora*, offer a platform for users to directly interact with each other. Reddit<sup>1</sup> is one of the largest online communities which contains a number of sub-communities (so called *subreddits*) that can be about almost anything. On this platform, several subreddits are specifically tailored to mental health-related topics, such as *depression*, *anxiety* or *alcoholism*. These *fora* offer a unique opportunity to study the way people describe or discuss their problems in their own voice.

One of the most challenging, and devastating, global mental health concerns is suicide. Suicidal behaviour includes any thoughts, plans or acts someone makes towards ending their life. In health care services, preventing death by suicide is a priority, but accurately predicting whether or not someone is at risk of committing suicide is difficult. Moreover, a large proportion of deaths by suicide occur in populations that have never been seen by health service providers.you could probably improve the section above...

Several online platforms are used for expressing suicidal thoughts and reaching out for support. On Reddit, the subreddit *SuicideWatch* currently<sup>2</sup> has almost 94k subscribers, and is

---

<sup>1</sup><http://reddit.com/>

<sup>2</sup>As of 27th June 2018

a moderated forum that is intended to offer peer support for people at risk of, or are worried about others', suicidal behaviour. The moderators take the message of peer support seriously, and are governed by guidelines that prohibits false promises, abuse, tough love and other clinically frowned upon methods of conversations<sup>3</sup>

As such it is valuable to understand what characteristics supportive communities like SuicideWatch have, and in what aspects such communities are similar or dis-similar from other casual subreddit conversations.

Recent studies have shown promising results in modeling and measuring signals and patterns in reddit communities related to mental health. For instance, statistical relations of mental health and depression communities with suicidal ideation have been studied [10, 11]. The authors explored linguistic and social characteristics that evaluate user's propensity to suicidal ideation. Approaches to classify reddit posts as related to certain mental health conditions have also been successfully developed, showing that there are certain characteristics specific to mental health-related topics in posts that can be automatically captured[16]. Furthermore, in a study focused on reddit posts related to anxiety, depression and post-traumatic stress disorder, the authors show that these online communities exhibit themes of supportive nature, e.g. gratitude for receiving emotional support[33]. Positive effects in participation in such fora have also been shown by improvements in members' written communication[32]. The supportive nature of comments in the SuicideWatch forum has also been studied by automatic identification and classification of helpful comments with promising results[23].

Most previous studies have aimed at studying the *content* of posts and their characteristics in relation to other posts. One important aspect of online communities is its supportive *function* — users turn to these platforms not only to express their thoughts and concerns, but also to receive support from the community. **More references to be added in the introduction. Also, we need to add something about the Online disinhibition effect somehow.**

To our knowledge, there are no studies that have specifically focused on modeling the supportive *nature* of online fora related to mental health. This work takes a macroscopic perspective, to quantitatively characterise and model the nature of supportive conversations. SuicideWatch is particularly interesting because of its purpose to offer peer support to people with suicidal thoughts, and also because of the complexity of this clinical construct.

Our aims in this study are to

- Understand similarities and differences between a Suicide watch conversation and a generic conversation using these abstraction.

---

<sup>3</sup><http://www.bbc.co.uk/newsbeat/article/35577626/social-media-and-suicide-what-its-like-being-a-moderator>

Terminology	stands for
<i>RP</i>	Root post which begins a new thread on a subreddit
<i>OP</i>	Original poster who posts the Root post for a thread
<i>SW</i>	The suicide watch Reddit
<i>FP</i>	Front page of Reddit.

Table 3.1 Notations and Terms.

- Study global properties of these conversations in comparison with control conversations.
- User network metrics to reason about global differences in terms of local interactions between users.

To model the network topology in an online community, we represent conversations in a forum using graph-based abstractions (users and replies) as described in Section 3.3.2. To measure global structure of these conversations, we user network topological metrics such as centrality: which measures importance of nodes in a network in terms of relaying information, branching factor: which measures how a conversation fans our over time, return distance: which measures how soon do users return back to the conversation and symmetric edges: which measures reciprocity of users in a conversation. To measure measure local interactions, we measure inter response times: which measure urgency of response to a message, semantic alignment between messages and local interaction motifs known as Triadic motifs : which gives an idea about how distinctive are interactions between subgroups of users.

## 3.2 Results

Particularly Suicide watch community consists of over 78k subscribers and reader, however is supported by mere 12 Moderators according to the latest count. The moderators are mainly present to prevent any kind of abuse, trolling or non-clinical or non-productive advices. These moderators do not have any form of formal training. However through several accounts they have confessed to learn through interactions and mentorship from more senior moderators on the site<sup>4</sup> All the moderators have been in that role for at least 3 years and the oldest goes as far as December 2008.

Our study is based on all conversations on SuicideWatch. We represent these conversations through networked abstractions as described in 3.3.2.

Through our analysis we find several discriminatory factors among Suicide watch conversations and generic front page conversation. We show that some of these factors are

---

<sup>4</sup><https://bbc.in/24rJYQH>

predictive of suicide watch conversations to a very high degree. We also show that certain properties of these conversations can be backed by sociological theories of real life support conversations.

### 3.2.1 Peculiarity of threads of Support

We begin by characterizing the two networked abstractions, namely Reply Graphs and Interaction graphs as described in Section 3.3.2. We do so by first comparing these two abstractions with a baseline control conversation threads using certain macroscopic network properties. We first compare the number of unique users per thread. The two communities exhibit considerable difference. SW Sub-Reddit has a median of 5 users per thread and a mean of 6.7 users and BL threads have a median of 25 users and mean of 50 unique users. So this off the bat indicates that Suicide watch conversations are more intimate and involve less participants. Authors tend to participate on a similar level, with a median participation per author of Suicide Watch to be 5 and for baseline conversations to be 3.

#### How urgent are user responses?

Understanding the inter message times can act as a good proxy for the urgency in a conversation. To understand how Suicide watch subreddit users responds to a *OP* compared to other sub-reddit threads on the frontpage, we calculate differences between the posting times between consecutive messages in a reply graph. Figure 3.1a shows comparison using CDFs of inter-message response times for SW and FP threads. It can be seen that SW *OP* are responded with the highest urgency amongst the 4, especially compared to either the *OP* or any other users or sub-reddits.

#### How symmetric are the interactions?

Despite signs of urgency and engagement, we ask the question: what percentage of conversations happening on these subreddits are symmetric in nature ? For this The median value for  $U_{sym}$  for SW is 20% where as for AS is 0%. This shows that SW subreddit engages in a lot more symmetric conversation than the baseline threads. If we define a set of users who engage in symmetric activity with the *OP*, it would be worth while to investigate how much of the total message activity on the thread is carried out by these set of symmetric users . To calculate this we find the fraction of messages on each thread written as part of this symmetric conversation. Figure 3.1b shows the trend. It can be see that SW threads contain a higher prevalence of symmetric message exchanges compared to the baseline

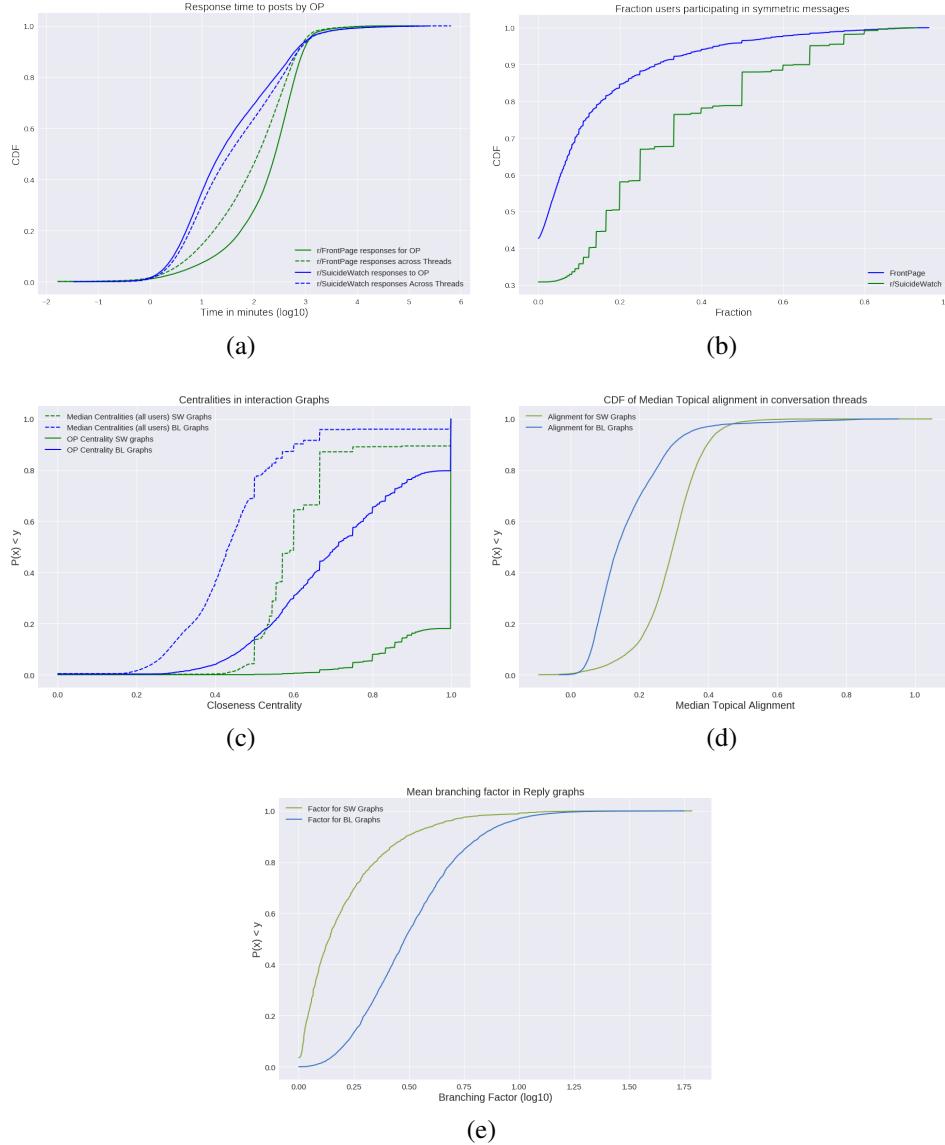


Fig. 3.1 Panel shows CDFs of different network metrics. Fig.3.1a shows the response time distributions, Fig.3.1b shows symmetrically engaged users, Fig.3.1d shows topical similarities across posts and 3.1e shows the branching factors of reply graphs.

Frontpage threads. This shows a higher engagement from the *OPs* side when participating in a supportive conversations

### How central are the users?

To understand how embedded is the *OP* in a conversation thread, we compare the betweenness centralities of *OPs* in the *SW* dataset with the baseline *FP* dataset. Betweenness centrality

is a good proxy of understanding how closely linked is a node with the rest of the network. When we calculate this metric for the user graphs we see that Suicide watch *OPs* tend to have highest centralities compared to generic *FP* threads both in terms of *OP* centrality as well as median centrality across all the users. The high centrality of *OPs* in *SW* conversations implies a high level of embeddedness as well as a *OP* centric approach by other participants in the conversation. The Figure 3.1c shows the Empirical CDFs of centralities.

### **How semantically aligned are the responses?**

We measure semantic alignment based on word embeddings of the source post and the reply post, at every edge of the reply graph. The detailed method of extracting semantic alignment along a post and its response is described in Section 3.3.3. Extracting such similarity metrics, we compare the trend in response text being in semantic alignment with the parent text in the reply graphs.

### **How branched do conversations become?**

Branching in a conversation thread could be either a sign of digression or a sign interestingness resulting in more people joining in. To measure this phenomena, use the reply graphs, which resemble a n-ary directed acyclic graph, to evaluate the branching factor. By using the method described in Section 3.3.3, we found that Suicide watch threads, tend to branch in a much less as compared to our baseline conversations. This implies that suicide watch threads tend to remain a one-on-one conversation with the *OP* albeit many such dialogues may emerge, and hence that explains the high centrality of the *OP* in all interaction graphs. If the helpers on a thread seldom interact with each other, the corresponding interaction graph will have the *OP* as the most central node.

### **3.2.2 Patterns in local interactions**

It is often a useful tool to express large interaction graphs, as the sum of local interactions between two or three nodes at a time. Such analysis is quite useful in expressing local structures in the graphs and has been used in several network analysis works. For this reason we use a method more commonly known as network motif analysis (described in Section 3.3.4) to understand triads, or groups of three nodes, and the patterns of edges that exist between them.

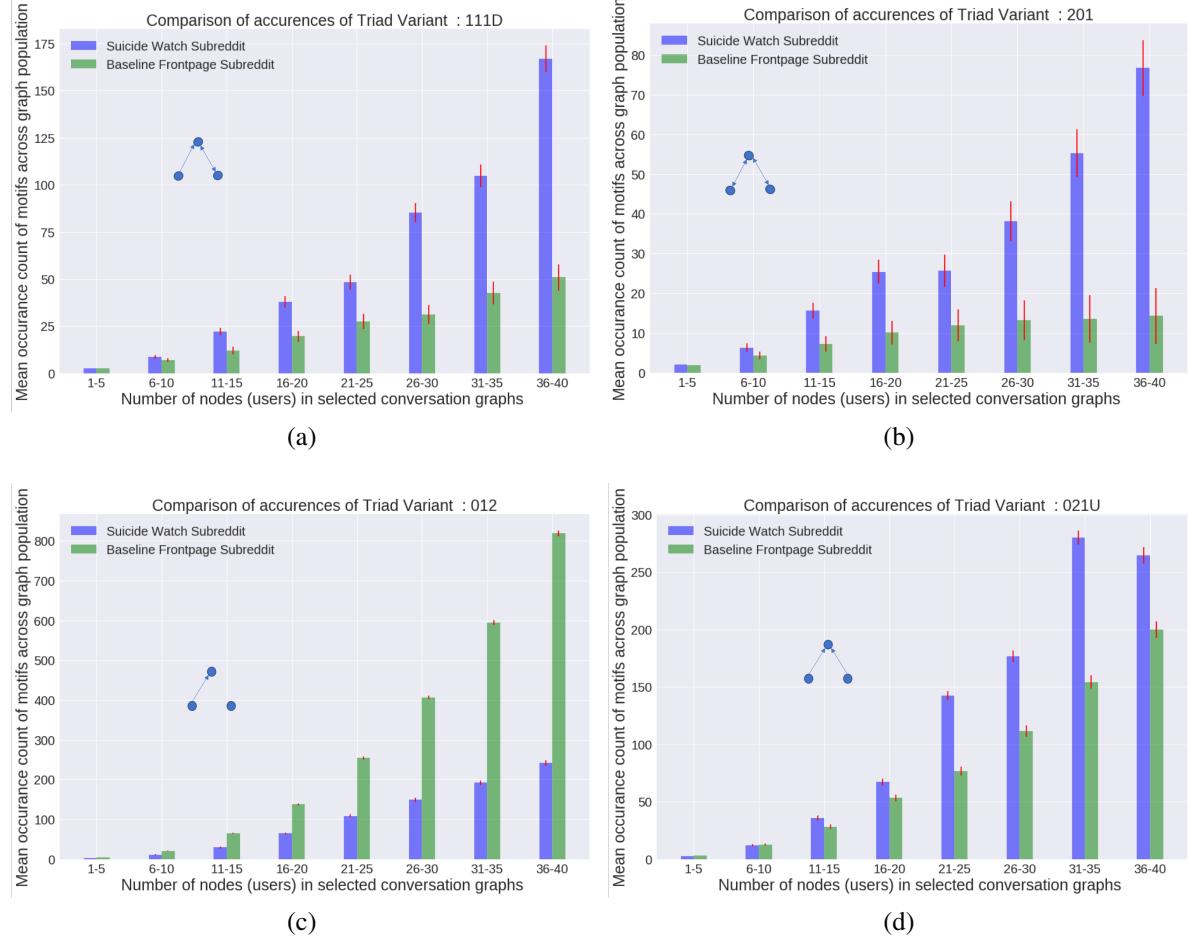


Fig. 3.2 This panel shows the statistical significance of the three over expressed and one under expressed triadic motif.

### 3.3 Methods

This section discusses the methodological devices used to extract insights from the fora data.

#### 3.3.1 Data

We build on dataset that was used in [16] where they analyze textual content for the root posts in a Subreddit called Suicide Watch<sup>5</sup>. The dataset contains a dump of 53 thousand posts from the suicide watch sub-reddit. However the dataset did not contain the threaded conversations for each thread. Reddit is a platform where a user can create a post on a sub-reddit, to which several members of a given sub-reddit can interact with. The array of interactions may range

<sup>5</sup><https://www.reddit.com/r/SuicideWatch/>

from simple up or down votes or posting at different hierarchy of the thread. This creates a hierarchical threaded structure of posts where the conversations are organized as threads of posts. To understand the deeper structure that is present in these posts, we crawl Reddit to get the threaded conversations by pursuing each conversation at arbitrary depth.<sup>6</sup>. This results in a dataset of over 50 thousand threads totaling to around 500,000 individual posts on those threads.

To baseline our work and compare theorized supportive nature of conversations with the broader community, we also crawl other reddit threads. To avoid any bias towards a particular type of subreddit, which have their own culture, we acquire roughly 50 thousand baseline posts which have been popular enough to land on the front page<sup>7</sup>. We crawl the Frontpage posts for 2 weeks accumulating over 50 thousand reddit threads in the process. The median amount of responses for a Suicide watch thread were 6 and for baseline Frontpage posts were 8. To understand the structure of these two forums, and find discriminating factors between a supportive community like suicide watch and a general thread on Reddit, we need to build abstractions of the thread.

### 3.3.2 Abstractions

To understand the dynamics of supportive conversations, we first need to formalize the abstraction of networked conversations. In case of forum based platforms where users interact in a nested dialogue fashion, and original poster or *OP* posts a start of a thread. This thread is then open for comments by all the community users. In case of Reddit, such a community is called a Subreddit, which is a moderated collection of users who subscribe to it. These users may post new threads onto the subreddit as far as the post follows the subreddit rules. Enforcement of these rules is the responsibility of the moderators. The user who starts a thread is called the Original Poster or **OP** and the headlining post which the *OP* begins with is called the Root Post or *RP*.

#### Reply Graphs

The first abstraction mimics directly the structure of conversation threads on Reddit. These abstractions are called Reply Graphs. We formulate a reply graph  $R\{P, E, W\}$  as a thread of multi-layered posts in a thread in response to the root post *RP* in the sub-reddit. Each graph  $R$  consists of posts  $P_i, P_j, i, j \in N$ , where  $N+1$  is the total number of responses in the thread and Edges  $E_{ij}$  such that and Edge  $E_{ij}$  exists iff post  $P_i$  was in response to post  $P_j$  in

---

<sup>6</sup>The code to crawl reddit for threads can be found at <https://github.com/sagarjoglekar/redditTools>

<sup>7</sup>The reddit front page algorithm is a combination of popularity and decay in popularity as a function of time. More can be found here <https://goo.gl/uVdHjn>

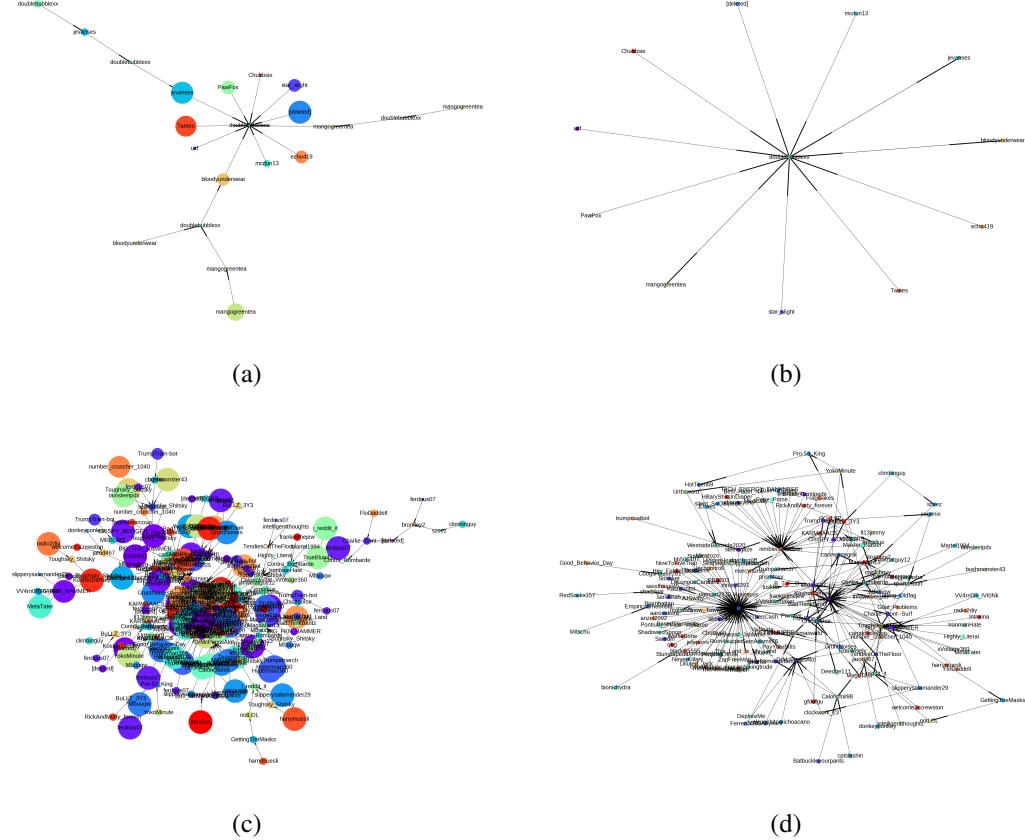


Fig. 3.3 Example UserGraphs and their corresponding Reply graphs, Figure 3.3b shows a random thread from the SW sub-reddit and 3.3a shows the corresponding reply graph that arises from the response structure of the same thread. In comparison we have Usergraph Fig 3.3d and its corresponding reply graph Fig 3.3c from one of the Front page threads

the hierarchy of responses. The weight of the edge  $E_{ij}$  is found by calculating the cosine similarity between topic vector  $T_i$  for post  $P_i$  and the topic vector  $T_j$  of post  $P_j$ . For a given dataset, the topic vectors are extracted using the model trained on that particular corpus ( $LDA_{BL}, LDA_{SW}$ ). This abstraction works well in modeling the conversational nature of these forums. For convenience of the reader, we present a couple of example pairs from SW and Frontpage baseline datasets in Figure 3.3

### Interaction Graphs

In this method, we represent each thread as a directed graph  $G\{V, E, W\}$  where  $V$  is the set of all users participating in a particular thread and  $E$  are the directed edges which correspond to interactions between two users  $V_i, V_j \in V$ . The weight of each directed edge  $E_{ij}$  corresponds

to the average of all the edges between  $V_i, V_j \in V$  in the corresponding reply graph  $R\{P, E, W\}$  as described above. This means that each reply graph is then mapped to a User graphs where the nodes are Users rather than posts. Another salient distinction between the two abstractions is that reply graphs resemble an n-ary tree and user graphs are directed cyclic graphs.

### 3.3.3 Macro and local metrics

The abstractions are used to extract certain structural metrics from the conversation threads. These metrics are then used to validate structural differences between supportive conversations and generic casual conversations from our baseline set.

#### Centrality

For this metric we use the User Graphs. Node centrality is a metric that measures how central a node is in a network. It directly reflects the importance of the node when it comes to membership of the shortest connecting paths between all the nodes in the graph. More formally, we use betweenness centrality of a node which is defined as Betweenness centrality of a node  $v$  is defined as

$$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

where  $\sigma_{st}(v)$  is the total number of shortest paths from node  $s$  to node  $t$  and  $\sigma_{st}$  is the number of those paths that pass through  $v$ . To understand whether the thread starters ( $OP$ ) have a special place in the network, we evaluate both  $OP$  centrality as well as median centrality across all the nodes in a User graph.

#### Symmetrical users

We define a symmetric user and a symmetric edges for user graphs. For a user  $V_i$  in the user graph  $G\{V, E, W\}$  as described in Section , a symmetric user is a user who interacts with  $V_o$  or the  $OP$  and receives a response back from the  $OP$ . We find the fraction

$$U_{sym} = \frac{\text{total number of symmetric users}}{\text{Total users in a thread}}$$

#### Urgency

To understand how Suicide watch subreddit users responds to the  $OP$  and each other, compared to other sub-reddit threads on the frontpage, we calculate differences between the

posting times between consecutive messages in a reply graph. We compute the median response times per thread, for *OP*'s posts and for all other posts.

### Branching Factor

Branching factor is a quantity that reflects the fan out of a conversation as it evolves. To measure this phenomena, we use the reply graphs, which resemble a n-ary directed acyclic graph, to evaluate the branching factor. The branching factor is formally described as

$$\tau = \frac{1}{|D|} \sum_{d \in D} \frac{1}{|N_d|} \sum_{n \in N_d} \text{InDeg}(n)$$

### Response distance

A user who interacts frequently with a thread, may contribute at different stages of hierarchy. Intuitively, for a one on one conversation to exist, the user needs to contribute back to the thread at alternate levels of hierarchy. To measure this phenomenon, we assume that a user  $U$  contributes in a reply graph at variable depths  $d_i \forall i \in [0, D]$ . We calculate the average difference between two consecutive contributions  $d_{avg}$

$$d_{avg} = \frac{\sum_{i=0}^D d_{i+1} - d_i}{D}$$

We calculate the values of  $d_{avg}$  for both *OP* and other users in the thread.

### Semantic similarity

We use a popular word embedding method called *Word2Vec* [28] which learns representations of a set of words from a corpus of text, which in our case is the text from Suicide Watch and baseline fora. These representations can be used to extract text embedding vectors for each post which belong to a  $N$  dimensional space  $R^N$ . These vectors are tested for their alignment using cosine distance in  $R^N$ , which corresponds to semantic similarity in the textual space. This method is quite popular and used in community based question answering[27], Medical semantic similarity [13] and other medical informatics applications[53].

#### 3.3.4 Structural metrics

Network motifs are local sub-networks between 2 or 3 nodes. Such local patterns are highly useful in quantifying local interactions and the resulting macro structure of the network[29]. They have been used in a variety of applications and networks, from economics [51] to

cellular protein-protein interaction networks [49]. Hence we calculate census of 16 distinct *triadic motif* i.e. possible subgraphs that could be formed between any three chosen nodes.

This probably needs to be briefly introduced/explained earlier in the manuscript? To understand relation of local structures in interaction networks within a conversation with its , we compare the quantity called *motif occurrence ratio*. After calculating motif census across the dataset[4], we select progressive subsets of graphs from both datasets with nodes  $n \in [k, k + 4] \forall k \in \{1, 6, 11, 16, 21, 26, 31, 36\}$ . This segmentation of the dataset is called binning, and it allows us to not only measure the differences in the census between the datasets, but also observe the trend as the size of the interaction graph becomes larger. We stop at 40 because beyond that, we start getting lower and lower number of graphs samples per bin.

We test each graph for the frequency of occurrence of the 16 possible triadic motifs as shown in Figure3.4a. We start by selecting the subset of graphs from Suicide watch and Frontpage belonging to a particular bin  $k$ . We then define the motif occurrence ratio as the fraction values for  $\frac{Y_{BL}}{K_{BL}}$  and  $\frac{Y_{SW}}{K_{SW}}$  for all the 16 motifs across all the chosen values of  $n$ .

## 3.4 Appendix

### 3.4.1 Triadic statistics for twitter conversations

#### 3.4.2

Replication of quantification of the topological metrics for twitter conversations for suicidal ideation and comparison with baseline threads. There are 5k threads for suicidal ideation and 6k for baseline. The baseline threads deal with discussion around westminster and manchester terrorist attacks in the uk.

#### 3.4.3 Network characteristics

Figure 3.7a shows the distribution of maximum depths across all Reply graphs for SW and Baseline subreddits. The SW threads depths have a median depth of 2 and mean of 4 compared to median depth of 2 for BL and a mean of 2.5. This shows that statistically the depths of Suicide watch and baseline graphs are quite similar.

Figure3.8 shows the CDF for the number of responses a Root post gets on a thread across the whole dataset. Figure 3.8 shows the CDF for number of comments per thread across the r/SuicideWatch subreddit dataset and the crawled frontpage subreddit.

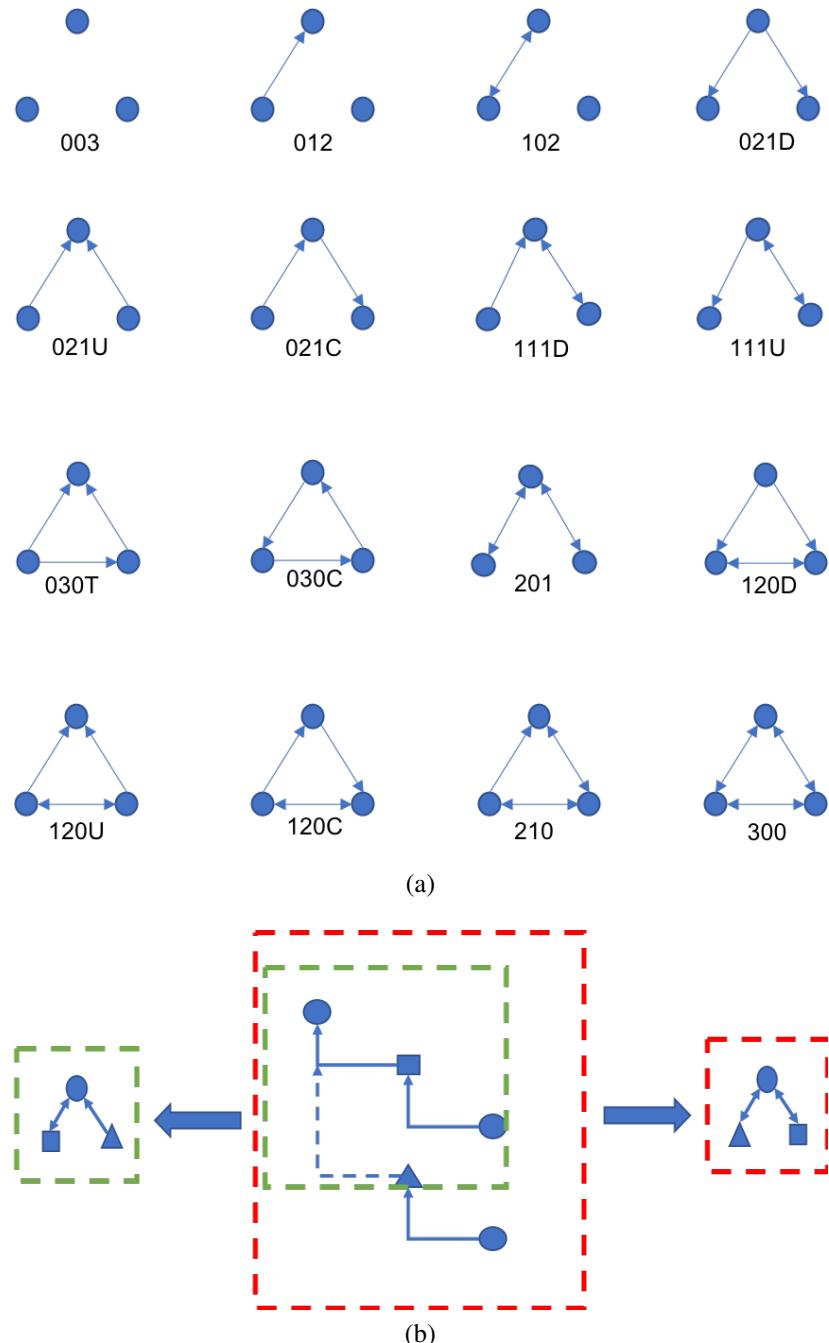


Fig. 3.4 Figure 3.4a shows the 16 different types of motifs that are looked for in the user graph data. Figure 3.4b shows how three unique users could produce different motifs. The three shapes represent different users and the dotted line means the message order is irrelevant.



Fig. 3.5 The figure shows comparison of occurrence ratios of 9 insignificant motifs. Blue traces are for Suicide watch and Green traces are for Baseline Front page threads

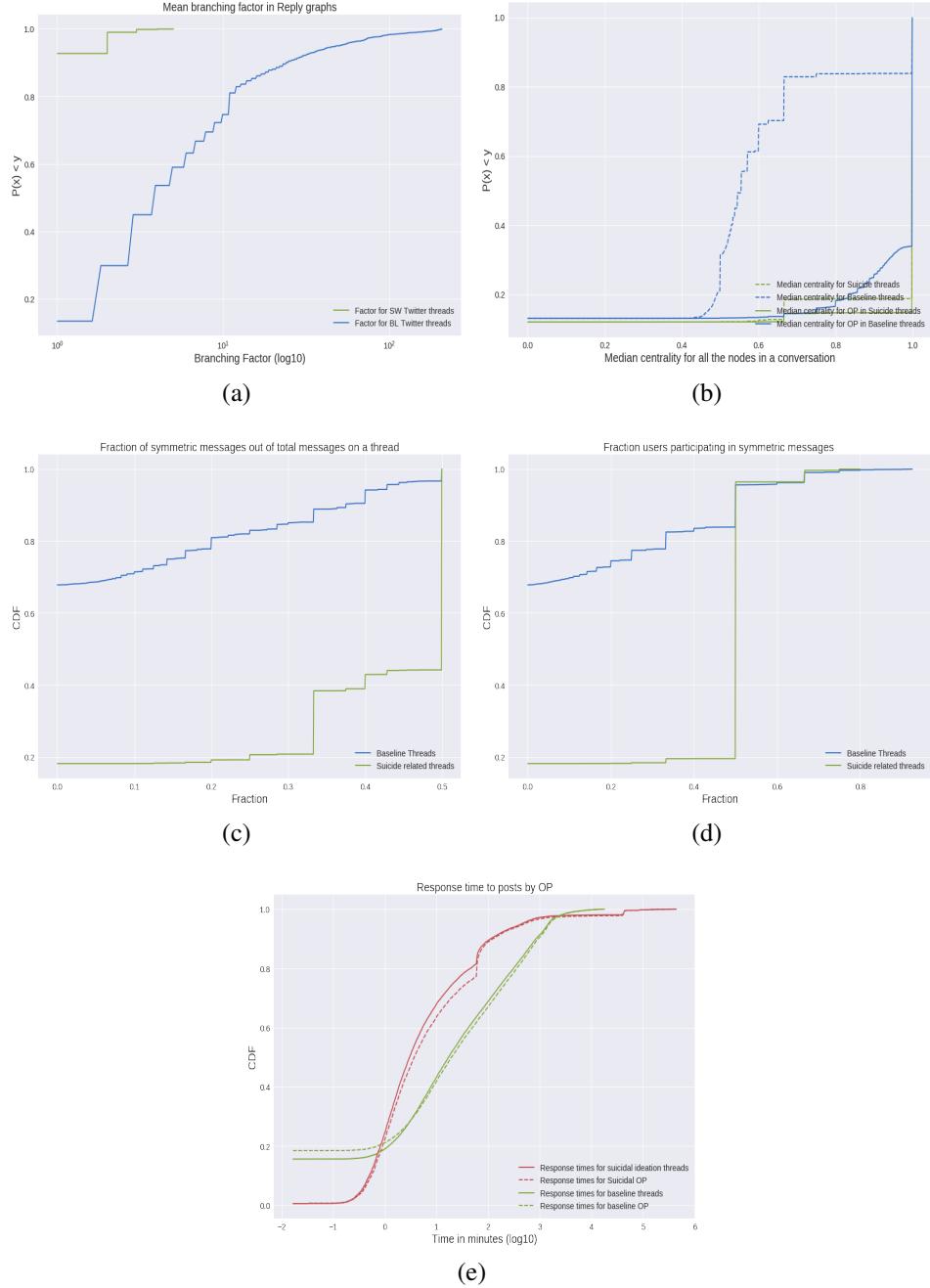


Fig. 3.6 Fig 3.6a shows the branching factor for twitter threads that talk about suicidal tendencies against baseline threads. Fig 3.6b shows the distribution of median centralities per thread, for both the twitter crawls. Fig 3.6c shows Distribution of symmetric messages in reply graphs for both datasets. Fig 3.6d shows the distributions for users participating in a symmetric conversation Fig 3.6e shows the distribution of reply urgency for suicide threads against baseline. The suicide median reponse time for suicide threads is 3 min as compared to 18 mins for non-suicide threads

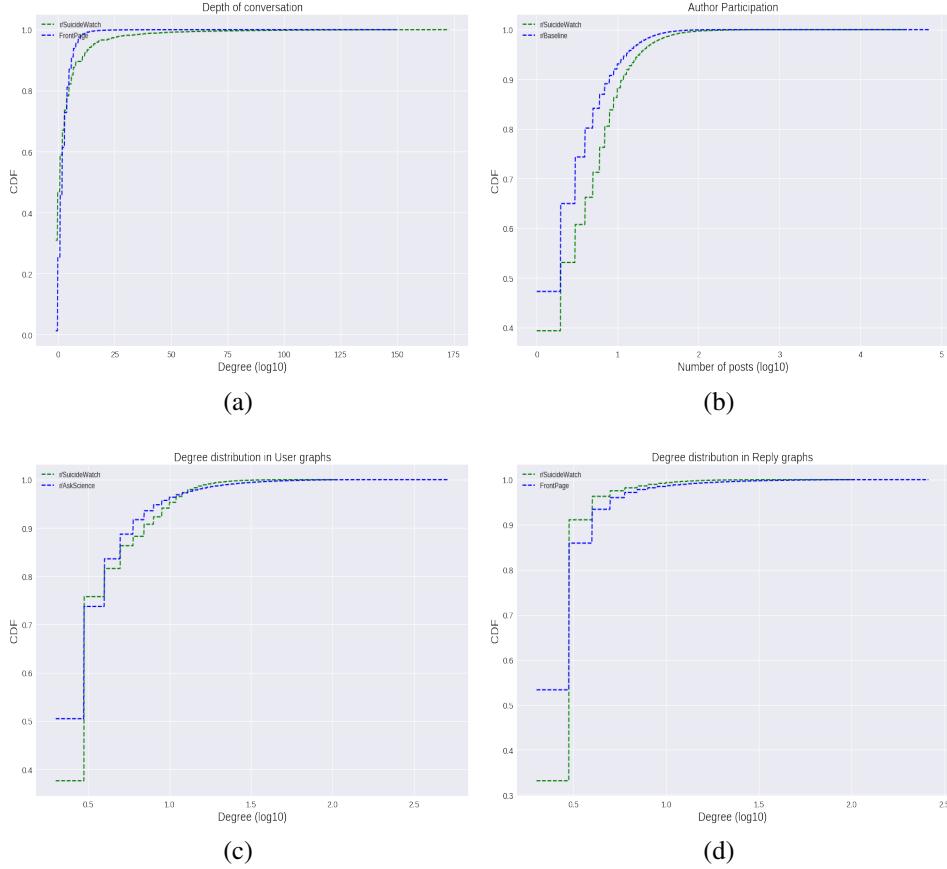


Fig. 3.7 Fig 3.7a shows the distribution of maximum depths of Reply Graphs for Subreddit r/SuicideWatch and the baseline Frontpage conversations. Fig 3.7b shows the distribution of unique authors per thread in the two datasets. Fig 3.7d shows Distribution of degrees for Reply Graphs, r/SuicideWatch and FrontPage. Fig 3.7c shows the degree distributions for the reply graphs

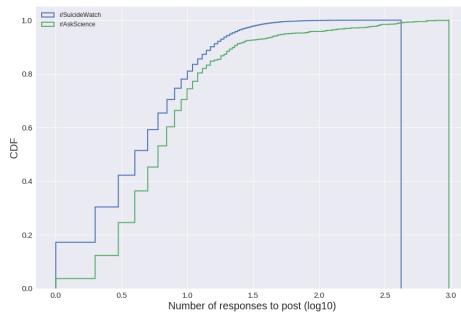


Fig. 3.8 Distribution of responses per thread on Subreddits r/SuicideWatch and Frontpage

# CHAPTER 4

## PERCEPTIONS IN REAL SPACES

---

There has been an explosive growth of deep learning technologies and their competency in the recent years, resulting in cross disciplinary use cases of deep learning enabled tools. In the area of computer vision and urban informatics, deep learning techniques have recently been used to predict whether urban scenes are likely to be considered beautiful, and it turns out that these techniques do so quite accurately. However, the technology falls short when it comes to generating actionable insights for AI assisted urban design. To support urban interventions, one needs to go beyond *predicting* beauty, and tackle the challenge of *recreating* beauty and *explaining* the predictors of beauty. Unfortunately, deep learning techniques have not been designed with that challenge in mind. Given their “black-box nature”, they cannot even explain why a scene has been predicted to be beautiful. To partly fix that, we propose a deep learning framework (which we name FaceLift) that is able to both *beautify* existing Google Street views and *explain* which urban elements make those transformed scenes beautiful. To quantitatively evaluate our framework, we cannot resort to any existing metric (as the research problem at hand has never been faced before) and need to formulate new ones. These new metrics should ideally capture the presence (or absence) of elements that make urban spaces great. Upon a review of the urban planning literature, we identify four main metrics: walkability, green, openness, and visual complexity. For all the four metrics, the beautified scenes meet the expectations set by the literature on what great spaces tend to be made of. The transformations and their explanations are also found to be very helpful in understanding interventions for beautification, which we validate using a 20-participant expert survey. These results suggest that, in the future, as our framework’s components are further researched and become better and more sophisticated, it is not hard to imagine technologies that will be able to accurately and efficiently support architects and planners in the design of the spaces we intuitively love.

## 4.1 Introduction

Whether a street is considered beautiful is subjective, yet research has shown that there are specific urban elements that are universally considered beautiful: from greenery, to small streets, to memorable spaces [1, 36, 38]. These elements are those that contribute to the creation of what the urban sociologist Jane Jacobs called ‘urban vitality’ [20].

Given that, it comes as no surprise that computer vision techniques can automatically analyse pictures of urban scenes and accurately determine the extent to which these scenes are, *on average*, considered beautiful. Deep learning has greatly contributed to increase these techniques’ accuracy [14].

However, urban planners and architects are interested in urban interventions and, as such, they wish to go beyond technologies that are only able to predict beauty scores. These interests stem from the fact that the spaces we live in can be linked with several aspects of human life such as mental health[41], inequality [38] or cultural shifts [18]. They often called for technologies that would make easier to recreate beauty in urban design [9]. Deep learning, by itself, is not fit for purpose. It is not meant to recreate beautiful scenes, not least because it cannot provide any explanation on why a scene is beautiful.

To partly fix that, we propose a deep learning framework (which we name FaceLift) that is able to both *generate* a beautiful scene (or, better, *beautify* an existing scene) and *explain* why that scene is beautiful. This opens up a possibility of using technology in urban planning efforts like decision making based of subjective opinions, participatory urban planning and promotion of restorative urban design such as green spaces and walkable areas. Through this work, we make two main contributions:

- We propose a deep learning framework that is able to learn whether Google Street views are beautiful or not, and that, based on that training, is able to both *beautify* existing views and *explain* which urban elements make these views beautiful (Section 4.3).
- We quantitatively evaluate whether the framework is able to actually produce beautified scenes (Section 4.4). We do so by proposing a family of four urban design metrics that we have formulated based on a thorough review of the literature in urban planning. For all these four metrics, the framework passes with flying colors: with minimal interventions, beautified scenes are twice as walkable as the original scenes, for example. Also, after building an interactive tool with “FaceLifted” scenes in Boston and presenting it to twenty experts in architecture, we found that the majority of them agreed on three main areas of our work’s impact: decision making, participatory urbanism, and promotion of restorative spaces among the general public.

For sake of brevity, we will use the term ‘Urban Scene’ through out the paper to address an arbitrary Google Street View image. The image is fetched from a particular latitude and longitude point on the map. In the rest of the paper we explore related literature across various tracks of urban perceptions and urban beauty in Section 4.2. We then describe in detail the Facelift framework in Section 4.3. The evaluation of the framework is described in detail in Section 4.4. We conclude by pointing out some limitations and biases that might well guide future work (Section 4.5).

## 4.2 Related Work

Previous work has focused on collecting ground truth data about how people perceive urban spaces, on predicting urban qualities from visual data, and on generating synthetic images that enhance a given quality (e.g., beauty).

**Ground truth of urban perceptions.** So far the most detailed studies of perceptions of urban environments and their visual appearance have relied on personal interviews and observation of city streets: for example, some researchers relied on annotations of video recordings by experts [39], while others have used participant ratings of simulated (rather than existing) street scenes [26]. The web has recently been used to survey a large number of individuals. Place Pulse is a website that asks a series of binary perception questions (such as ‘Which place looks safer [between the two]?’) across a large number of geo-tagged images [38]. In a similar way, Quercia *et al.* collected pairwise judgments about the extent to which urban scenes are considered quiet, beautiful and happy [36]. They were then able to analyze the scenes together with their ratings using image-processing tools, and found that the amount of greenery in any given scene was associated with all three attributes and that cars and fortress-like buildings were associated with sadness. Taken all together, their results pointed in the same direction: urban elements that hinder social interactions were undesirable, while elements that increase interactions were the ones that should be integrated by urban planners to retrofit cities for greater happiness.

**Deep learning and the city.** Computer vision techniques have increasingly become more sophisticated. Deep learning techniques, in particular, have been recently used to accurately predict urban beauty [14, 42], urban change [30], and even crime [12]. These works also did some interesting analysis of the data to understand how safety, depression, beauty and other such dimensions are perceived across urban spaces. [14] also utilized deep learning methods to train models capable of comparing two urban images for their perception values in terms

Symbol	Meaning
$I_i$	Original urban scene
$Y$	Set of annotation classes for urban scenes (e.g., beautiful, ugly)
$y_i$	Annotation class in $Y$ (e.g., beautiful)
$\hat{I}_j$	Template scene (synthetic image)
$I'$	Target Image
$C$	Beauty Classifier

Table 4.1 Notations

of beauty et.al. However even these works did not dive into the reasoning aspect of these models.

**Generative models.** Deep learning has recently been used not only to analyze existing images but also to generate new ones. In the past couple of years, there have been papers which exploit generative version of neural nets to delve into the aspects of explainability. Nguyen *et al.* [31] used generative networks to create a natural-looking image that maximizes a specific neuron. In theory, the resulting image is the one that “best activates” the neuron under consideration (e.g., that associated with urban beauty). In practice, it is still a synthetic template that needs further processing to look realistic.

To sum up, a lot of work has gone into collecting ground truth data about how people tend to perceive urban spaces, and into building accurate predictions models of urban qualities. However, little work has gone into models that generate realistic urban scenes and that offer human-interpretable explanations of what they generate.

### 4.3 FaceLift Framework

The goal of FaceLift is to take as input a geo-located urban scene and give as output its transformed (beautified) version. To that end, it performs in five steps:

- **Curating urban scenes** It is common knowledge that deep learning systems need immense amount of data. In this first step we try to develop a sound framework for curating and augmenting annotated images, on which the model could be trained.

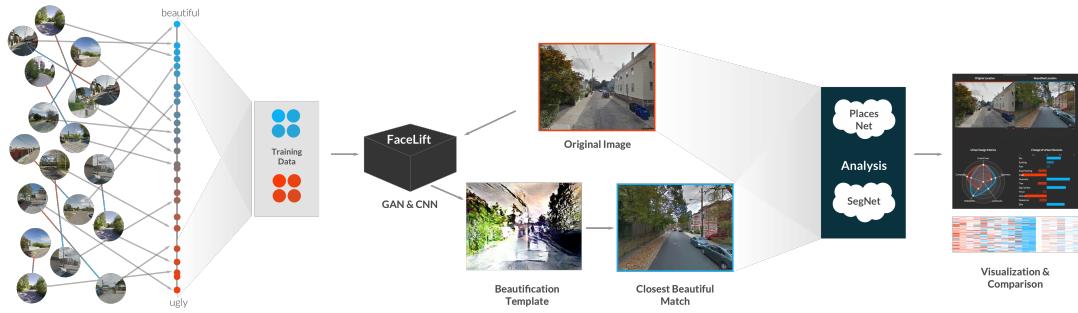


Fig. 4.1 A simplistic end to end illustration of the FaceLift framework.

- **Training a beauty classifier** To generate beauty, you first need a reliable model that could learn the representation of beauty. To achieve this, we train a deep learning model that could distinguish beautiful urban scenes from non-beautiful urban scenes.
- **Generating a synthetic beautified scene** Based on the learned representation of beauty, we train a Generative model which could augment the beauty of an input urban scene.
- **Retrieving a realistic beautified scene** as showcased in Figure 4.1, the generated images are representations of beautified input urban scene in a latent space. This latent representation needs to be transformed back to a realistic looking image, using retrieval.
- **Identifying the urban elements characterizing the beautified scene** In the final step, the framework explains changes introduced in the transformation process in terms of literature-driven urban design metrics, and quantifies these changes as metrics for urban beauty.

## Step 1 Curating Urban Scenes

To begin with, we need highly curated training data with labels reflecting urban beauty. We start with the Place Pulse dataset that contains 100k Google Street Views across 56 cities around the world [14]. These scenes are labeled in terms of whether the corresponding places are likely to be perceived beautiful, depressing, rich, and safe. We focus only on those scenes that are labeled in terms of beauty and that have at least three judgments. This leave us with roughly 20,000 scenes. To transform judgments into beauty scores, we use the TrueSkill algorithm [17], which gives us a way of partitioning the scenes into two sets (Figure 4.2): one containing beautiful scenes, and the other containing ugly scenes. The resulting set of scenes

Augmentation	Accuracy (Percentage)
None	63
Rotation	68
Rotation + Translation	64
Rotation + Conservative Translation	73.5

Table 4.2 Percentage accuracy for our beauty classifier trained on differently augmented sets of urban scenes.

is too small for training any deep learning module without avoiding over-fitting though. As such, we need to augment such a set.

We do so in two ways. First, we feed each scene’s location into the Google Streetview API to obtain the snapshots of the same location at different camera angles (i.e., at  $\theta \in -30^\circ, -15^\circ, 15^\circ, 30^\circ$ ). However, the resulting dataset is still too small for robust training. Therefore, again, we feed each scene’s location into the Google Streetview API, but now we do so to obtain other scenes at distance  $d \in \{10, 20, 40, 60\}$  meters. This will greatly expand our set of scenes, but it might do so at the price of introducing scenes whose beauty scores have little to do with the original scene’s. To fix that, we take only the scenes that are *similar* to the original one (we call this way of augmenting “conservative translation”). To compute the similarity between a pair of scenes, we represent the two scenes with visual features derived from the FC7 layer of PlacesNet and compute the similarity between the two corresponding feature vectors [52]. For all scenes at increasing distance  $d \in \{10, 20, 40, 60\}$  meters, we take only those whose similarity scores with the original scene is above a threshold. In a conservative fashion, we choose that threshold to be the median similarity between rotated and original scenes (those of the first augmentation step).

To make sure this additional augmentation has not introduced any unwanted noise, we consider two sets of scenes: one containing those that have been taken during this last step, i.e. the one with high similarity to the original scenes (*taken-set*), and the other containing those that have been filtered away (*filtered-set*). Each scene is then scored with PlacesNet [52] and is represented with the five most confident scene labels. We then aggregate labels at set level, by computing each label’s frequency on the *taken-set* and on the *filtered-set*. Finally, we characterize each label’s propensity to be correctly augmented as:  $\text{prone}(\text{label}) = \text{fr}(\text{label}, \text{taken-set}) - \text{fr}(\text{label}, \text{filtered-set})$ . This reflects the extent to which a scene with a given label is prone to be augmented or not. From Figure 4.4, we find that, as one would expect, scenes that contain highways, fields and bridges can be augmented at increasing distances while still showing resemblances to the original scene; by contrast, scenes that contain gardens, residential neighborhoods , plazas, and skyscrapers cannot be easily augmented, as they are often found in high density parts of the city, where there is tremendous diversity within short distances.

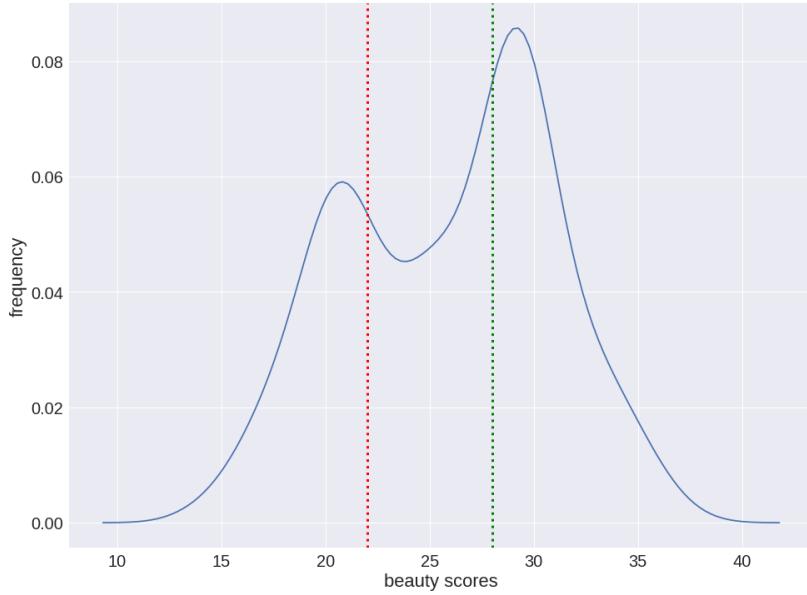


Fig. 4.2 Frequency distribution of beauty scores. The red and green lines represent the thresholds below and above which images are considered ugly and beautiful. Conservatively, images in between are discarded.

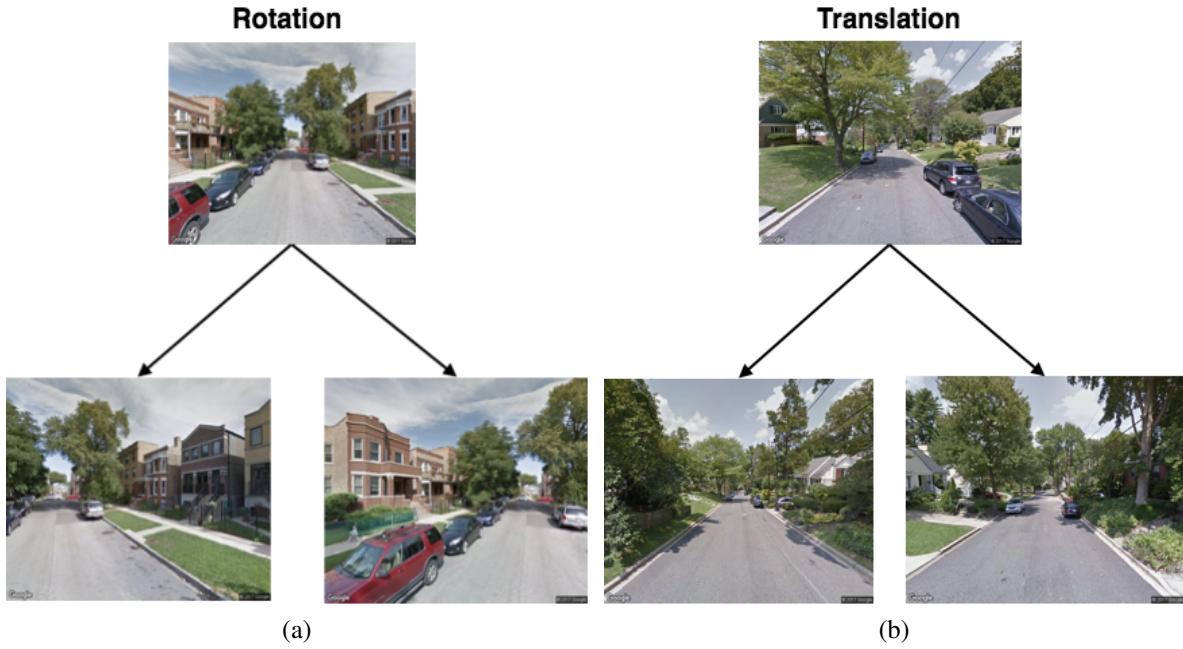


Fig. 4.3 Two types of augmentation: (a) rotation of the Street Views camera (based on rotation); and (b) exploration of scenes at increasing distances (based on translation).

## Step 2 Training a beauty classifier

Having this highly curated set of labeled urban scenes, we are now ready to train classifier  $C$  with labels reflecting our beauty assessments. We use the CaffeNet architecture, a modified

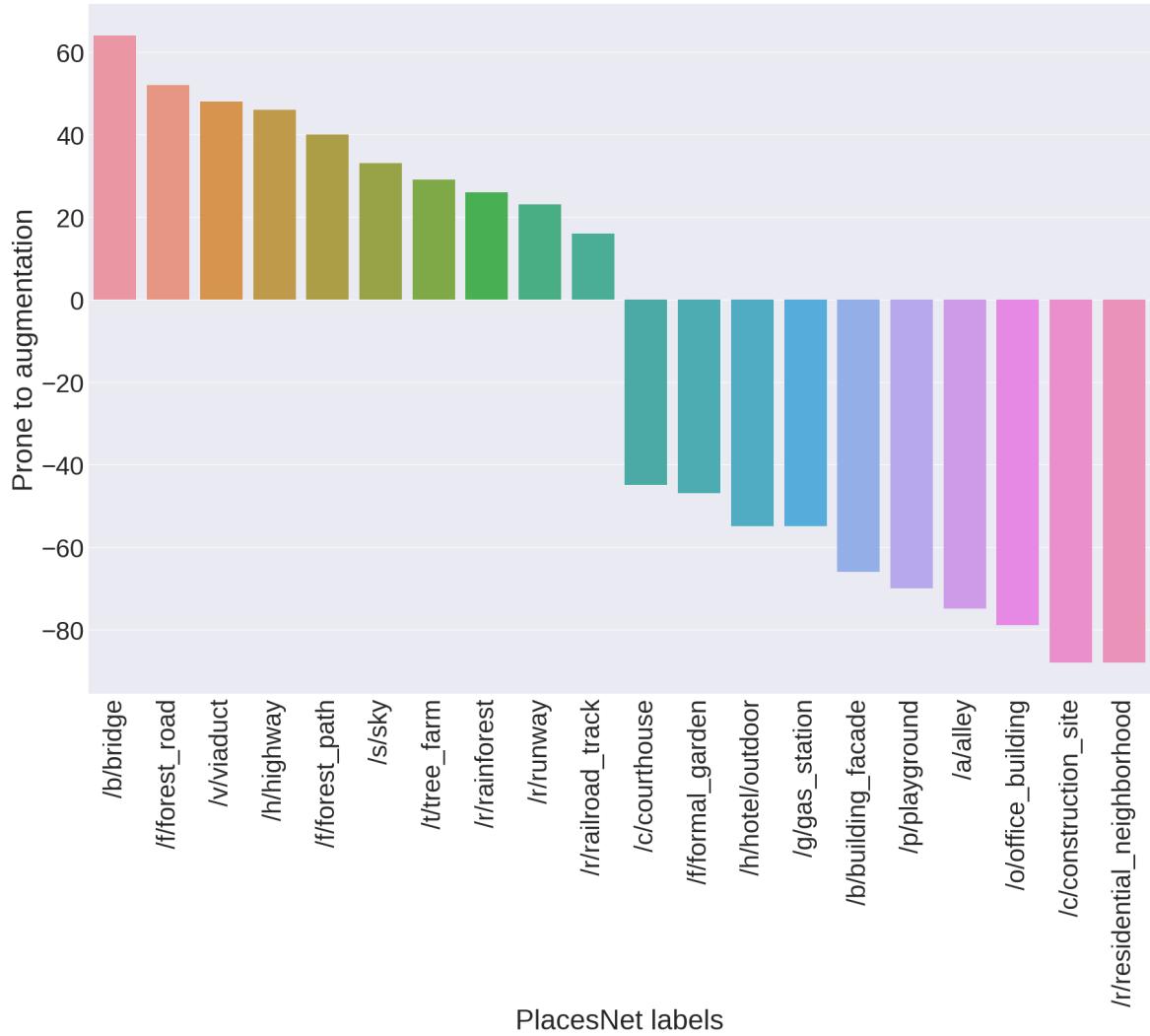


Fig. 4.4 The types of scene that have greater propensity to be correctly augmented with similar scenes at increasing distances.

version of AlexNet [24, 47]. The training is done on a 70% split of the data, and the testing on the remaining 30%. All this is done on increasingly augmented sets of data. We start from our 20k images and progressively augment them with the snapshots obtained with the 5-angle camera rotations, and then with the exploration of scenes at increasing distance  $d \in \{10, 20, 40, 60\}$  meters. The idea behind data augmentation is that accuracy would increase with it. Indeed it does (Table 4.2): it goes from 63% on the set of original scenes to as much as 73.5% on the set of fully augmented scenes, which is a notable increase in accuracy for such classes of classification tasks. As a baseline, we compare with the models trained by Dubey et.al in [14] on the same seed data that we use for our pipeline. They report that their models perform at 70% accuracy in the task of picking a beautiful image amongst

any two given images. Albeit the set-up of our model is not to compare two images but just to classify a particular image in a binary class, this baseline shows that our model is showing a comparable performance in beauty classification.

### Step 3 Generating a synthetic beautified scene

Having this trained classifier at hand, we can then build a generator of synthetic beautified scenes. This is a model that, given the two classes ugly  $y_i$  and beautiful  $y_j$ , transforms any original scene  $I_i$  of class  $y_i$  (e.g., ugly scene) into template scene  $\hat{I}_j$  that maximizes class  $y_j$  (e.g., beautified template scene).

More specifically, given an input image  $I_i$  known to be of class  $y_i$  (e.g., ugly), our technique outputs  $\hat{I}_j$ , which is a more beautiful version of it (e.g.,  $I_i$  is morphed towards the average representation of a beautiful scene) while preserving  $I_i$ 's details. The technique does so using the “Deep Generator Network for Activation Maximization” (*DGN-AM*) [31]. Given an input image  $I_i$ , *DGN-AM* iteratively re-calculates the color of  $I_i$ 's pixels in a way the output image  $\hat{I}_j$  both maximizes the activation of neuron  $y_j$  (e.g., the “beauty neuron”) and looks “photo realistic”, which is done by conditioning the maximization to an “image prior”. This is equivalent to finding the feature vector  $f$  that maximizes the following expression:

$$\hat{I}_j = G(f) : \arg \max_f (C_j(G(f)) - \lambda ||f||) \quad (4.1)$$

where:

- $G(f)$  is the image synthetically generated from the candidate feature vector  $f$ ;
- $C_j(G(f))$  is the activation value of neuron  $y_j$  in the scene classifier  $C$  (the value to be maximized);
- $\lambda$  is a  $L_2$  regularization term.

Here the initialization of  $f$  is key. If  $f$  were to be initialized with random noise, then the resulting  $G(f)$  would be the average representation of category  $y_j$  (of, e.g., beauty). Instead, since  $f$  is initialized with  $I_i$ , then the resulting  $G(f)$  is  $I_i$ 's version “morphed to become more beautiful”.

### Step 4 Returning a realistic beautified scene

We now have template scene  $\hat{I}_j$  (which is a synthetic beautified version of original scene  $I_i$ ) and need to retrieve a realistic looking version of it. We do so by: *i*) representing each of our

original scenes in Step 1 (including  $\hat{I}_j$ ) as a 4096 dimensional feature vector derived from the FC7 layer of the PlacesNet [52]; *ii*) computing the distance (as  $L_2$  Norm) between  $\hat{I}_j$ 's feature vector and each of the original scene's feature vector; and *iii*) selecting the original scene most similar (smaller distance) to  $\hat{I}_j$ . This results into the selection of the beautified scene  $I_j$ .

## Step 5 Identifying characterizing urban elements

Since original scene  $I_i$  and beautified scene  $I_j$  are real scenes and we make sure that they maintain the same structural characteristics (e.g., point of view, layout), we can easily compare them in terms of presence or absence of SegNet's and PlacesNet's labels. That is, we can determine how the original scene and its beautified version differ in terms of urban design elements. This step required us to develop metrics inspired from urban design literature, to quantify the changes in elements. A detailed description of the characterization and evaluation would follow in Section 4.4

## 4.4 Evaluation

The goal of FaceLift is to transform existing urban scenes into versions that: *i*) people perceive more beautiful; *ii*) contain urban elements typical of great urban spaces; *iii*) are easy to interpret; and *iv*) architects and urban planners find useful. To ascertain whether FaceLift meets that composite goal, we answer the following questions next:

**Q1** Do individuals perceive “FaceLifted” scenes to be beautiful?

**Q2** Does our framework produce scenes that possess urban elements typical of great spaces?

**Q3** Which urban elements are mostly associated with beautiful scenes?

**Q4** Do architects and urban planners find FaceLift useful?

### Q1 People's perceptions of beautified scenes

To ascertain whether FaceLifted scenes are perceived by individuals as they are supposed to, we run a crowd-sourcing experiment on Amazon Mechanical Turk. We randomly select 200 scenes, 100 beautiful and 100 ugly (taken at the bottom 10 and top 10 percentiles of the Trueskill's score distribution of Figure 4.2). Our framework then transforms each ugly scene into its beautified version, and each beautiful scene into its corresponding ‘uglified’. These scenes are arranged into pairs, each of which contains the original scene and its beautified or

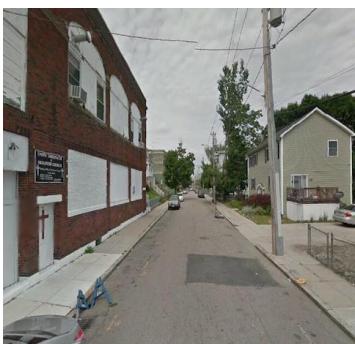
Original ( $I_i$ )	Latent Beauty representation ( $\hat{I}_j$ )	Beautified ( $I_j$ )
		
		
		
		
		

Table 4.3 The table showcases examples of the “FaceLifting” process. It is worth observing

uglified version. On Mechanical Turk, we only select verified masters for our crowd-sourcing workers (those with an approval rate above 90% during the past 30 days), pay them \$0.1 per task, and ask each of them to choose the beautiful scene for given pairs. We make sure to have at least 3 votes for each scene pair. Overall, our workers end up selecting the scenes that are actually beautiful 77.5% of the times, suggesting that FaceLifted scenes are correctly perceived most of the times.

## Q2 Are beautified scenes great urban spaces?

To answer that question, we need to understand what makes a space great. After a careful review of the urban planning literature, we identify four factors [1, 15] (summarized in Table 4.4): great places mainly tend to be walkable, offer greenery, feel cozy, and be visually rich.

Metric	Description
Walkability	Walkable streets increase the social capital of a place, and they appeal to the exploring nature of the human psyche [15, 35, 45].
Green Spaces	The presence of greenery has repeatedly been found to impact people's well being [1]. Under certain conditions, it could also promote social interactions [36]. This suggest that not all greenery has to be considered in the same way though: dense forests or unkempt greens might well have a negative impact [20].
Privacy-Openness	A sense of privacy (as opposed to a sense of openness) impacts a place's perception [15].
Visual Complexity	Visual complexity is a measure of how diverse a urban scene is in terms of design materials, textures, and objects [15].

Table 4.4 Urban Design Metrics

To automatically extract visual cues related to these four factors, we select 500 ugly scenes and 500 beautiful ones at random, transform them into their opposite aesthetic qualities (i.e., ugly ones are beautified, and beautiful ones are ‘uglified’), and compare which urban elements related to the four factors distinguish uglified scenes from beautified ones.

We extract labels from each of our 1,000 scenes using two image classifiers. First, using PlacesNet [52], we label each of our scenes according to a classification containing 205 labels (reflecting, for example, landmarks, natural elements), and retain the five labels with highest confidence scores for the scene. Second, using Segnet [2], we label each of our scenes

according to a classification containing 12 labels. Segnet is trained on dash-cam images, and classifies each scene pixel with one of these twelve labels: road, sky, trees, buildings, poles, signage, pedestrians, vehicles, bicycles, pavement, fences, and road markings.

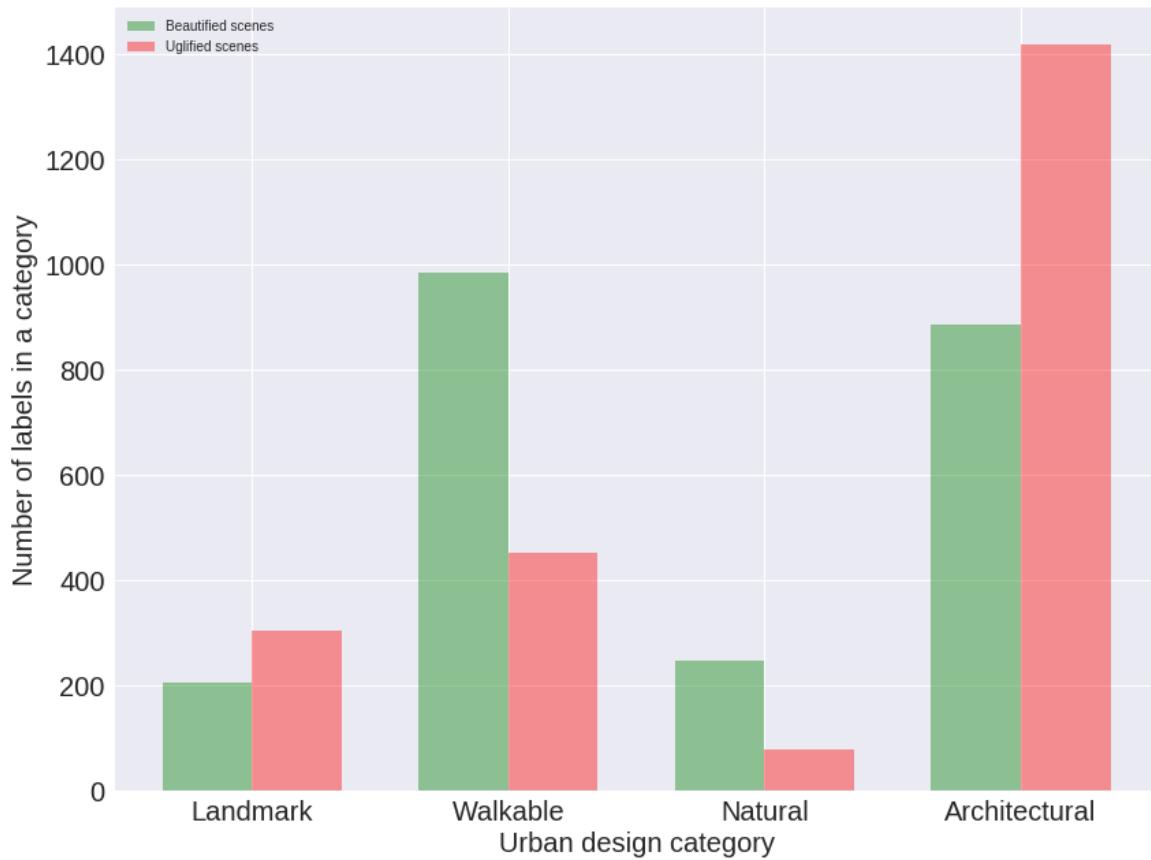


Fig. 4.5 Number of labels in specific urban design categories (on the *x*-axis) found in beautified scenes as opposed to those found in uglified scenes.

Having these two ways of labeling scenes, we can now test whether the expectations set by the literature describing metrics of great urban spaces (Table 4.4) are met in the FaceLifted scenes.

*H1 Beautified scenes tend to be walkable.* We manually select only the PlacesNet labels that are related to walkability. These labels include, for example, *abbey*, *plaza*, *courtyard*, *garden*, *picnic area*, and *park*. To test hypothesis *H1*, we count the number of walkability-related labels found in beautified scenes as opposed to those found in uglified scenes (Figure 4.5): the former contain twice as many walkability labels than the latter. We then determine which types of scenes are associated with beauty (Figure 4.6). Unsurprisingly, beautified scenes

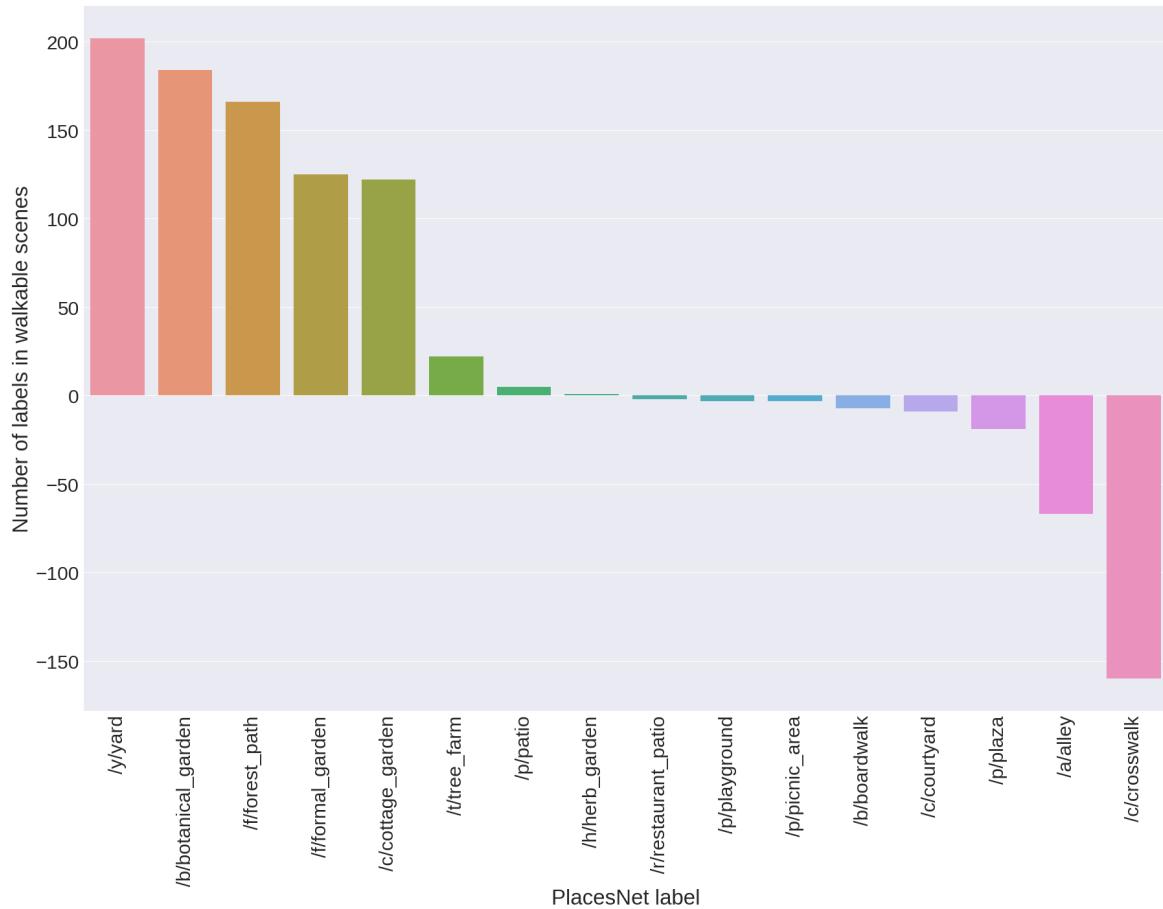


Fig. 4.6 Count of specific walkability-related labels (on the *x*-axis) found in beautified scenes minus the count of the same labels found in uglified scenes.

tend to show gardens, yards, and small paths. By contrast, uglified ones tend to show built environment features such as shop fronts and broad roads.

*H2 Beautified scenes tend to offer green spaces.* We manually select only the PlacesNet labels that are related to greenery. These labels include, for example, *fields*, *pasture*, *forest*, *ocean*, and *beach*. Then, in our 1,000 scenes, to test hypothesis *H2*, we count the number of nature-related labels found in beautified scenes as opposed to those found in uglified scenes (Figure 4.5): the former contain more than twice as many nature-related labels than the latter. To test this hypothesis further, we compute the fraction of ‘tree’ pixels (using SegNet’s label ‘tree’) in beautified and uglified scenes, and find that beautification adds 32% of tree pixels, while uglification removes 17% of them.

*H3 Beautified scenes tend to feel private and ‘cozy’.* To test hypothesis *H3*, we count the

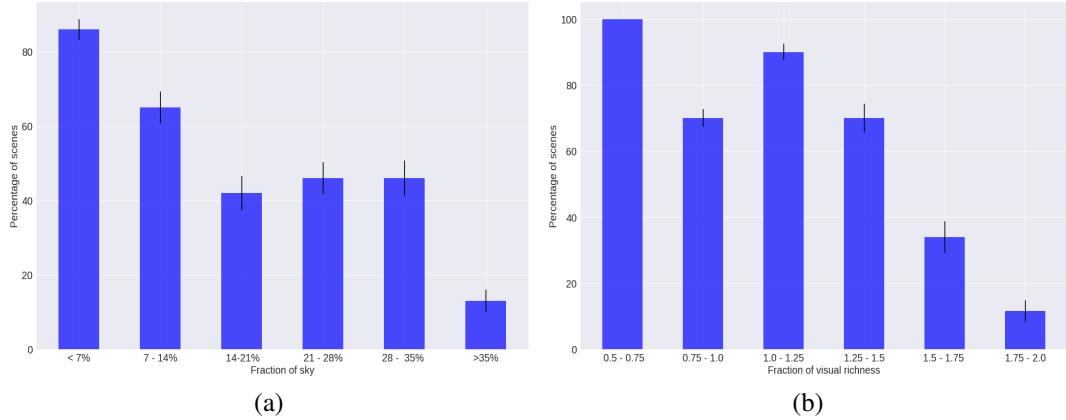


Fig. 4.7 The percentage of scenes (y-axis): (a) having an increasing presence of sky (on the x-axis); and (b) having an increasing level of visual richness (on the x-axis). The error bars represent standard errors obtained by random re-sampling of the data for 500 iterations.

fraction of pixels that Segnet labeled as ‘sky’ and show the results in a bin plot in Figure 4.7a: the x-axis has six bins (each of which represents a given range of sky fraction), and the y-axis shows the percentage of beautified vs. uglified scenes that fall into each bin. Beautified scenes tend to be cozier (lower sky presence) than the corresponding original scenes.

*H4 Beautified scenes tend to be visually rich.* To quantify to which extent scenes are visually rich, we measure their visual complexity [15] as the amount of disorder in terms of distribution of (Segnet) urban elements in the scene:

$$H(X) = - \sum p(i) \log p(i) \quad (4.2)$$

where  $i$  is the  $i^{th}$  Segnet’s label. The total number of labels is twelve. The higher  $H(X)$ , the higher the scene’s entropy, that is, the higher the scene’s complexity. To test hypothesis *H4*, we show the percentage of scenes that fall into a complexity bin (Figure 4.7b): beautified scenes are of low to medium complexity, while uglified ones are of high complexity.

### Q3 Urban elements of beautified scenes

To determine which urban elements are the best predictors of urban beauty and the extent to which they are so, we run a logistic regression, and, to ease interpretation, we do so on one pair of predictors at the time:

$$Pr(\text{beautiful}) = \text{logit}^{-1}(\alpha + \beta_1 * V_1 + \beta_2 * V_2 + \beta_3 * V_1.V_2) \quad (4.3)$$

Pair of urban elements	$\beta_1$	$\beta_2$	$\beta_3$	Error Rate (Percentage)
Buildings - Trees	-0.032	0.084	0.005	12.7
Sky - Buildings	-0.08	-0.11	0.064	14.4
Roads - Vehicles	-0.015	-0.05	0.023	40.6
Sky - Trees	0.03	0.11	-0.012	12.8
Roads - Trees	0.04	0.10	-0.031	13.5
Roads - Buildings	-0.05	-0.097	0.04	20.2

Table 4.5 Coefficients of logistic regressions run on one pair of predictors at the time.

where  $V1$  is the fraction of the scene’s pixels marked with one Segnet’s label, say, “buildings” (over the total number of pixels), and  $V2$  is the fraction of the scene’s pixels marked with another label, say, “trees”. The result consists of three beta coefficients:  $\beta_1$  reflects  $V1$ ’s contribution in predicting beauty,  $\beta_2$  reflects  $V2$ ’s contribution, and  $\beta_3$  is the interaction effect, that is, it reflects the contribution of the dependency of  $V1$  and  $V2$  in predicting beauty. We run logistic regressions on the five factors that have been found to be most predictive of urban beauty [1, 15, 36], and show the results in Table 4.5.

Since we are using logistic regressions, the quantitative interpretation of the beta coefficients is eased by the “divide by 4 rule” [48]: we can take  $\beta$  coefficients and “divide them by 4 to get an upper bound of the predictive difference corresponding to a unit difference” in beauty [48]. For example, take the results in the first row of Table 4.5. In the model  $Pr(\text{beautiful}) = \text{logit}^{-1}(\alpha - 0.032 \cdot \text{buildings} + 0.084 \cdot \text{trees} + 0.005 \cdot \text{buildings} \cdot \text{trees})$ , we can divide  $-0.032/4$  to get  $-0.008$ : a difference of 1 in the fraction of pixels being buildings corresponds to no more than a 0.8% *negative* difference in the probability of the scene being beautiful. In a similar way, a difference of 1 in the fraction of pixels being trees corresponds to no more than a 0.021% *positive* difference in the probability of the scene being beautiful. By considering the remaining results in Table 4.5, we find that, across all pairwise comparisons, trees is the most positive element associated with beauty, while roads and buildings are the most negative ones. Since these results go in the direction one would expect, one might conclude that the scenes beautified by our framework are in line with previous literature, adding further external validity to our work.

## Q4 Do architects and urban planners find it useful?

To ascertain whether practitioners find FaceLift potentially useful, we built an interactive map of the city of Boston in which, for selected points, we showed pairs of urban scenes before/after beautification (Figure 4.8). We then sent that map along with a survey to 20 experts in architecture, urban planning, and data visualization around the world. The experts had to complete tasks in which they rated FaceLift based on how well it supports decision

Use case	Definitely Not	Probably Not	Probably	Very Probably	Definitely
Decision Making	4.8%	9.5%	38%	28.6%	19%
Participatory Urban Planning	0%	4.8%	52.4%	23.8%	19%
Promote Green Cities	4.8%	0%	47.6%	19%	28.6%

Table 4.6 Urban experts polled about the extent to which an interactive map of “FaceLifted” scenes promotes: (a) decision making; (b) citizen participation in urban planning; and (c) promotion of green cities



Fig. 4.8 Interactive map of FaceLifted scenes in Boston.

making, participatory urbanism, and promotion of green spaces among the general public. The results are shown in Table 4.6 according to our experts, the tool can very probably supports

decision making, probably support participatory urbanism, and definitely promote green spaces. These results are qualitatively supported by our experts' comments, which include: “*The maps reveal patterns that might not otherwise be apparent*”, “*The tool helps focusing on parameters to identify beauty in the city while exploring it*”, and “*The metrics are nice. It made me think more about beautiful places needing a combination of criteria, rather than a high score on one or two dimensions. It made me realize that these criteria are probably spatially correlated*”.

## 4.5 Conclusion

FaceLift is a transparent framework that beautifies existing urban scenes. This translates into two main technical advancements. First, FaceLift is able to generate realistic scenes as opposed to existing approaches based on Generative Adversarial Networks whose final transformations are quite coarse as they still take the form of synthetic templates. Second, it augments the deep learning black-box with a module that offers explanations on what has been transformed, making that box more transparent.

There are still important limitations though. One is data bias. The framework is as good as its training data, and more work has to go into collecting reliable ground truth of human perceptions. This data should ideally be stratified according to the people's characteristics that impact their perceptions. The other main limitation is that generative models are hard to control, and more work has to go into offering principled ways of fine-tuning the generative process.

Despite these limitations, FaceLift has the potential to support urban interventions in scalable and replicable ways: it can be applied to the scale of an entire city, and that can be replicated in other cities. The advantage of shifting the focus of research away from predictive analytics towards urban interventions is that people could be part of discussions on works of architecture more than they are nowadays. To turn existing spaces into something more beautiful, that will still be the duty of architecture. Yet, with technologies similar to FaceLift more readily integrated in the architecture discussions, the complex job of recreating restorative spaces in an increasingly urbanized world will be greatly simplified. After all, “we delight in complexity to which genius have lent an appearance of simplicity.” [9] In the context of future work, that genius is represented by the future technologies that we will contribute to build to deal with the complexity of our cities.

# CHAPTER 5

## PERCEPTION AND GENERATIVE MODELS

---



## REFERENCES

---

- [1] Alexander, C., Ishikawa, S., Silverstein, M., i Ramió, J. R., Jacobson, M., and Fiksdahl-King, I. (1977). *A pattern language*. Gustavo Gili.
- [2] Badrinarayanan, V., Kendall, A., and Cipolla, R. (2015). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *arXiv preprint arXiv:1511.00561*.
- [3] Bakhshi, S., Shamma, D. A., and Gilbert, E. (2014). Faces engage us: Photos with faces attract more likes and comments on instagram. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 965–974. ACM.
- [4] Batagelj, V. and Mrvar, A. (2001). A subquadratic triad census algorithm for large sparse networks with small maximum degree. *Social Networks*, 23(3):237–243.
- [5] Borges, M. (2004). What can kant teach us about emotions? *The Journal of Philosophy*, 101(3):140–158.
- [6] Cecchi, L., Liccardi, G., Pellegrino, F., and Sofia, M. (2012). 2 social networks: A new source of psychological stress or a way to enhance self-esteem? negative and positive implications in bronchial asthma. *Journal of Investigational Allergology and Clinical Immunology*, 22(6):402.
- [7] Crawford, M. (2015). *The World Beyond Your Head: How to Flourish in an Age of Distraction*. penguin publishing.
- [8] Davenport, T. H. and Beck, J. C. (2001). *The attention economy: Understanding the new currency of business*. Harvard Business Press.
- [9] De Botton, A. (2008). *The Architecture of Happiness*. Vintage Series. Knopf Doubleday Publishing Group.
- [10] De Choudhury, M. and De, S. (2014). Mental Health Discourse on reddit: Self-Disclosure, Social Support, and Anonymity. *Proceedings of the Eight International AAAI Conference on Weblogs and Social Media*, pages 71–80.
- [11] De Choudhury, M., Kiciman, E., Dredze, M., Coppersmith, G., and Kumar, M. (2016). Discovering shifts to suicidal ideation from mental health content in social media. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI ’16, pages 2098–2110, New York, NY, USA. ACM.

- [12] De Nadai, M., Vieriu, R. L., Zen, G., Dragicevic, S., Naik, N., Caraviello, M., Hidalgo, C. A., Sebe, N., and Lepri, B. (2016). Are Safer Looking Neighborhoods More Lively?: A Multimodal Investigation into Urban Life. In *Proceedings of the ACM on Multimedia Conference (MM)*.
- [13] De Vine, L., Zuccon, G., Koopman, B., Sitbon, L., and Bruza, P. (2014). Medical semantic similarity with a neural language model. In *Proceedings of the 23rd ACM international conference on conference on information and knowledge management*, pages 1819–1822. ACM.
- [14] Dubey, A., Naik, N., Parikh, D., Raskar, R., and Hidalgo, C. A. (2016). Deep learning the city: Quantifying urban perception at a global scale. *arXiv preprint arXiv:1608.01769*.
- [15] Ewing, R. and Clemente, O. (2013). *Measuring urban design: Metrics for livable places*. Island Press.
- [16] Gkotsis, G. (2017). Characterisation of mental health conditions in social media using informed deep learning. *Scientific Reports*, 7:45141.
- [17] Herbrich, R., Minka, T., and Graepel, T. (2007). Trueskill™: a bayesian skill rating system. In *Advances in neural information processing systems*, pages 569–576.
- [18] Hristova, D., Aiello, L. M., and Quercia, D. (2018). The new urban success: How culture pays. *Frontiers in Physics*, 6:27.
- [19] Hwang, K. O., Ottenbacher, A. J., Green, A. P., Cannon-Diehl, M. R., Richardson, O., Bernstam, E. V., and Thomas, E. J. (2010). Social support in an internet weight loss community. *International journal of medical informatics*, 79(1):5–13.
- [20] Jacobs, J. (1961). *The Death and Life of Great American Cities*. Random House.
- [21] Joglekar, S., Sastry, N., and Redi, M. (2017). Like at first sight: Understanding user engagement with the world of microvideos. In *International Conference on Social Informatics*, pages 237–256. Springer.
- [22] Kant, I. (1987). *Critique of judgment*. Hackett Publishing.
- [23] Kavuluru, R., Ramos-Morales, M., Holaday, T., Williams, A. G., Haye, L., and Cerel, J. (2016). Classification of helpful comments on online suicide watch forums. In *Proceedings of the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics, BCB '16*, pages 32–40, New York, NY, USA. ACM.
- [24] Krizhevsky, A. et al. (2012). Imagenet classification with deep convolutional neural networks. In *NIPS*, pages 1097–1105.
- [25] Kummervold, P. E., Gammon, D., Bergvik, S., Johnsen, J.-A. K., Hasvold, T., and Rosenvinge, J. H. (2002). Social support in a wired world: use of online mental health forums in norway. *Nordic journal of psychiatry*, 56(1):59–65.
- [26] Lindal, P. J. and Hartig, T. (2012). Architectural variation, building height, and the restorative quality of urban residential streetscapes. *Journal of Environmental Psychology*.

- [27] Mihaylov, T. and Nakov, P. (2016). Semanticz at semeval-2016 task 3: Ranking relevant answers in community question answering using semantic similarity based on fine-tuned word embeddings. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 879–886.
- [28] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119.
- [29] Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. (2002). Network motifs: simple building blocks of complex networks. *Science*, 298(5594):824–827.
- [30] Naik, N., Kominers, S. D., Raskar, R., Glaeser, E. L., and Hidalgo, C. A. (2017). Computer vision uncovers predictors of physical urban change. *Proceedings of the National Academy of Sciences*, 114(29):7571–7576.
- [31] Nguyen, A., Dosovitskiy, A., Yosinski, J., Brox, T., and Clune, J. (2016). Synthesizing the preferred inputs for neurons in neural networks via deep generator networks. In *Advances in Neural Information Processing Systems*, pages 3387–3395.
- [32] Park, A. and Conway, M. (2018). Harnessing reddit to understand the written-communication challenges experienced by individuals with mental health disorders: Analysis of texts from mental health communities. *J Med Internet Res*, 20(4):e121.
- [33] Park, A., Conway, M., and Chen, A. T. (2018). Examining thematic similarity, difference, and membership in three online mental health communities from reddit: a text mining and visualization approach. *Computers in human behavior*, 78:98–112.
- [34] Perlovsky, L. (2014). Aesthetic emotions, what are their cognitive functions? *Frontiers in Psychology*, 5:98.
- [35] Quercia, D., Aiello, L. M., Schifanella, R., and Davies, A. (2015). The Digital Life of Walkable Streets. In *Proceedings of the 24th ACM Conference on World Wide Web (WWW)*, pages 875–884.
- [36] Quercia, D., O’Hare, N. K., and Cramer, H. (2014). Aesthetic capital: what makes london look beautiful, quiet, and happy? In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, pages 945–955. ACM.
- [37] Rowley, J. (2007). The wisdom hierarchy: representations of the dikw hierarchy. *Journal of information science*, 33(2):163–180.
- [38] Salesses, P., Schechtner, K., and Hidalgo, C. A. (2013). The collaborative image of the city: mapping the inequality of urban perception. *PloS one*, 8(7):e68400.
- [39] Sampson, R. J. and Raudenbush, S. W. (2004). Seeing Disorder: Neighborhood Stigma and the Social Construction of Broken Windows. *Social Psychology Quarterly*, 67(4).
- [40] Schupp, H. T., Flaisch, T., Stockburger, J., and Junghöfer, M. (2006). Emotion and attention: event-related brain potential studies. *Progress in brain research*, 156:31–51.

- [41] Seresinhe, C. I., Preis, T., and Moat, H. S. (2015). Quantifying the impact of scenic environments on health. *Scientific reports*, 5:16899.
- [42] Seresinhe, C. I., Preis, T., and Moat, H. S. (2017). Using deep learning to quantify the beauty of outdoor places. *Royal Society open science*, 4(7):170170.
- [43] Soat, M. (2015). Social media triggers a dopamine high. *Marketing News*, 49(11):20–21.
- [44] Song, Y., Dixon, S., and Pearce, M. (2012). A survey of music recommendation systems and future perspectives. In *9th International Symposium on Computer Music Modeling and Retrieval*, volume 4.
- [45] Speck, J. (2012). Walkable City: How Downtown Can Save America, One Step at a Time. In *Farrar, Straus and Giroux*.
- [46] Squire, M. (2015). "should we move to stack overflow?" measuring the utility of social media for developer support. In *2015 IEEE/ACM 37th IEEE International Conference on Software Engineering*, volume 2, pages 219–228. IEEE.
- [47] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9.
- [48] Vaughn, B. K. (2008). Data analysis using regression and multilevel/hierarchical models, by gelman, a., & hill, j. *Journal of Educational Measurement*, 45(1):94–97.
- [49] Yeger-Lotem, E., Sattath, S., Kashtan, N., Itzkovitz, S., Milo, R., Pinter, R. Y., Alon, U., and Margalit, H. (2004). Network motifs in integrated cellular networks of transcription–regulation and protein–protein interaction. *Proceedings of the National Academy of Sciences*, 101(16):5934–5939.
- [50] Zadra, J. R. and Clore, G. L. (2011). Emotion and perception: The role of affective information. *Wiley interdisciplinary reviews: cognitive science*, 2(6):676–685.
- [51] Zhang, X., Shao, S., Stanley, H. E., and Havlin, S. (2014). Dynamic motifs in socio-economic networks. *EPL (Europhysics Letters)*, 108(5):58001.
- [52] Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., and Oliva, A. (2014). Learning deep features for scene recognition using places database. In *Advances in neural information processing systems*, pages 487–495.
- [53] Zhu, Y., Yan, E., and Wang, F. (2017). Semantic relatedness and similarity of biomedical terms: examining the effects of recency, size, and section of biomedical publications on the performance of word2vec. *BMC medical informatics and decision making*, 17(1):95.