A GRAPH ANALYSIS OF THE IFTTT SERVICE

Andrew S. T. Lee

Empirical Network Analysis Project
Social Network Analysis Course by Lada Adamic

Abstract

This study presents the results of an analysis of a user-created tasking network from an online service called IFTTT. IFTTT is a brokering service that provides a Web interface for registered users to arrange automated tasks between various services and device functions. Users can create useful tasks and often share these with other registered users which adopt these so-called *recipes* for personal use. In this study we discuss and summarize key results from public data gathered from IFTTT and we apply several graph algorithms to reveal characteristic usage patterns. We finally conclude on the study and propose other possibilities of study in converging networks of users and Internet-enabled devices.

I. INTRODUCTION

IFTTT is an abbreviation of *If This Then That* – this phrase encapsulates succinctly the primary function of the service. IFTTT is essentially a brokering service and "user configurable switchboard" that joins up useful functionality exposed by an online service or device feature to other services and devices [1]. One common use case is the knowledge of whether it will rain today (obtained from a weather reporting service) is conveyed to a user's smartphone as a timely visual alert (using a mobile phone's OS notification service). There exist other useful pairwise combinations of functions offered by the same or different service endpoints. IFTTT refers to a service endpoint as a *channel* and channels[1] are the basic building block of the service [2]. To name but a few, IFTTT channels include online social networks (e.g. Facebook), blogs (e.g. Tumblr), domestic sensors (e.g. Nest thermostat) and ubiquitous email and SMS services. A channel has its own *trigger* and *action* interface and users can create an automated task that connects one channel's trigger interface to another channel's action interface. This pairwise combination of trigger with action interfaces is known as a *recipe* [3].

The same service provider may offer one or more channels and if possible recipes might use the same channel for action and trigger. A user must register an account before personal recipes can be created. A newly created recipe is private to the user and can be shared with other users by publishing the recipe to a searchable directory. Before a recipe can be created or adopted from the published index, each of the recipe's constituent channels must be activated – this entails the user granting IFTTT service permission to access user information associated with an account linked to the channel or permission to trigger functions of a channel. To enable triggers or actions on a smartphone, the user is also required to install a native mobile client that has access to on-device functionality.

User-created recipes can be interpreted collectively as a single tasking network where each node represents a channel. We will analyze the properties of the resultant graph in Section IV. To date there exists a pool of 144

[1] The terms *channels* and *services* will be used synonymously hereafter

different channels[2] from which users can create their own recipes. There exist a total of 149537 recipes published from 73568 user accounts. Furthermore some user accounts may be associated with one or more private recipes – these represent an unknown number of unpublished personal recipes. For the purpose of this study we assume that the data obtained from published recipes are representative of usage patterns in both public and private lists. Moreover, disregarding the actual content of data conveyed from trigger to action endpoint, we envisage that the published data will betray popular relationships between channels. We assume that there exists an innate tendency for frequent users of the IFTTT service to share recipes that could be useful to other members.

II. DATA COLLECTION

Data for published IFTTT recipes were gathered from a Python script that polled periodically for a list of recipes from the main recipe search page [3]. This Web "scraping" task was distributed over several runs so as not to overload the site. A total of 15040 indexed Web pages were obtained which as of the final date of collection (5th November 2014) represented the entire set of public recipes. Another script was used to parse each page and extract the following details of each recipe:

- The recipe's trigger channel
- The recipe's action channel
- The user that created the recipe
- *Adds*: this variable counts the number of times a user has included the recipe in their personal list
- *Favorites*: this variable counts the number of times a user has *favorited* the recipe

There exist duplicates of recipes in the collected data as a recipe can appear in one or more lists on the search page. Each of these lists represent a particular curated set of recipes and are categorized as *Trending*, *Featured*, *New* and *All Time*. The script skips duplicate recipes and extracted data is pre-processed before detailed analysis with Gephi graph visualization [4] and R language [5].

III. RECIPE USE CASES

IFTTT users have created a diverse collection of use cases that automate some task as mediated by IFTTT's brokering infrastructure and IFTTT mobile app. IFTTT recipes can be divided roughly into the following three groups

1. Recipes that automate the transfer of personal data between channels
2. Recipes that periodically monitor for or triggered by certain events from one channel and signal such events onto another channel
3. Recipes that bind a command interface to a particular action

Table 1 lists example activities for each of the aforementioned group of recipes:

Group 1	Group 2	Group 3
• Photo transfer from smartphone to an online service • Transfer personal data to a cloud-based storage service • Save emails or Web pages to a cloud-based knowledge management service • Log some personal activity to a service	• Weather and news alerts, email updates • Device sensor reports (domestic or industrial) • Social network alerts • Location alerts • Reminders • Run simple automated tasks	• Speech based commands to automate household or online tasks

Table 1: Typical activities in ad-hoc classification of recipes

[2] *MixRadio* channel was added during the writing of this report increasing the total number of channels to 145

The activities listed in Table 1 are non-exhaustive and new forms of interaction will be invented or existing ones made more accessible as new services and APIs are exposed to consumers and to the wider service-product ecosystem. Recipes typically use two different channels and we model recipes and channels as a graph in Section IV. However, recipes can also use the same channel as both a trigger and action – recipes that re-organize content into new collections or some event-triggered actions fall into this category.

Table 2 shows the top 10 most popular recipes in terms of *adds* – this list suggest that most users of the IFTTT service are automating tasks between popular online services and their personal smartphones. We shall see in the next section that there also exists recipes that utilize the functions of an emerging class of networked devices and sensors and these recipes signal usage patterns that will become more prevalent as more devices are connected to the Internet.

Description	*Recipe ID*	*Trigger channel*	*Action channel*	*Adds*	*Favourites*
Backup my contacts to a Google Spreadsheet	102384	iOS Contacts	Google Drive	110639	4376
Rain tomorrow? Get an Android Notification	165139	Weather	Android Notifications	88697	2895
Get all the Updates to IFTTT via email!	8363	IFTTT	Email	80054	622
Email me my new iPhone photos	103371	iOS Photos	Email	79645	1633
If your Facebook profile picture changes then update your Twitter profile picture too	8981	Facebook	Twitter	70638	1777
Download Facebook Photos you're tagged in to Dropbox	15	Facebook	Dropbox	62768	1707
Rain tomorrow? Get an iOS Notification	153667	Weather	iOS Notifications	60287	2379
Mute my Android device when I get to the office & turn on vibrate	164966	Android Location	Android Device	53005	4010
Post your Instagram pictures as native Twitter pictures	103249	Instagram	Twitter	49977	3145
Save my iOS photos to Dropbox	103376	iOS Photos	Dropbox	41595	1048

Table 2: Top 10 most popular IFTTT recipes in ascending order of the number of adds.

IV. GRAPH OF IFTTT CHANNELS

We construct a *channel graph* G of available IFTTT channel to facilitate analysis of the types of recipes users are creating and to assess the popularity of certain recipes and channels. We also wish to ascertain patterns of usage that emerge from the popularity of certain channels.

Each node $v_i \in G$ represents a IFTTT channel and a directed edge (s, t) is added if there exists a recipe that connects a trigger channel s with action channel t[3]. For the directed graph G, we assign a weight $k_{st} = \max\left(|R_{st}|, \sum_{r=1}^{|R_{st}|} a_r\right)$ to each edge (s, t) where R_{st} is the set of all user created recipes incident on (s, t) and a_r is the number of adds for recipe $r \in R_{st}$. k_{st} represent the popularity of recipes incident on edge (s, t) and the minimum value for k_{st} will always be 1 as there must be at least one published recipe with no adds or at most one add. The graph is generated from collected data and imported into Gephi in GML format. Using Gephi we have applied several graph algorithms to the channel graph and present the results in the following sections. We use Gephi's *ForceAtlas 2* layout algorithm[4] for visualizing G and we tune settings to ensure nodes are distributed uniformly on the plotted space for clarity of the layout.

A. In-Degree and Out-Degree

The non-weighted in-degree for an arbitrary channel measures the number of unique recipes whose action endpoint is incident on the channel. Similarly, non-weighted out-degree measures the number of unique recipes whose trigger endpoint is incident on the channel. The non-weighted degrees measures the diversity of recipes in which a channel is associated with. The weighted in-degree and out-degree measures the popularity of the channel as an action or trigger respectively. Section VI will visualize the *inner core* of channels were most user recipes participate in.

[3] Note that self-loops can exist for those recipes that engage the same channel. These are omitted from the graph visualization.

[4] Settings for *ForceAtlas 2*: edge weight influence = 0.3, "prevent overlap" setting, scaling = 180.0. A lower edge weight influence is used to ensure a wide spread of nodes given the large range of edge weights. Manual tweaking of node drawing positions was also conducted for aesthetic reasons.

B. Community Detection

We wish to identify why some recipes are aligned to particular channels and whether there are strong underlying affinities between certain groups of channels. A community detection algorithm [6] was used to determine the modularity of 0.393 for graph G considering edge weights. Figure 1 depicts a rendering of graph G that highlights 5 distinct communities discovered by the algorithm and Figure 2(c) depicts the most popular recipes in each community. Each community is color coded for ease of identification and for visualization purposes. For each community, the top 5 most popular trigger channels, action channels and recipes are tabulated in Table 4.

The *green community* (Figure 2(a)) is the largest in terms of node and edge count and encompass recipes for household devices and services that play a role in the nascent *Internet of Things* [7]. Typical activities in this community include:

- Domestic devices and sensors (e.g. Nest thermostat) that communicate status to SMS, phone calls, iOS/Android notifications and other online services.
- Push of monitored events and status to SMS, phone calls, devices, iOS/Android notifications and other push brokers.
- Monitoring of a variety of events that include location data, dates, times, weather, news, social "pings" (e.g. Yo), incoming SMS and other types of service alerts.

The *magenta community* (Figure 2(b)) comprises scheduling of events in Google Calendar and life logging activities that collectively belong to the broader *quantified self* movement [8] which include a wider quantification of assets. This community encompasses the following main activities:

- Monitoring of signals (measurements from fitness devices, life events and personal assets)
- Scheduling events to and receiving alerts from Google Calendar
- Google Drive as action channel i.e. store content from other channels
- iOS Location as trigger channel i.e. record personal geolocation to other channels
- LinkedIn as both an action and trigger channel

The membership of Google Drive channel indicate that users are using this online storage service primarily for archiving photos and other forms of personal content from other channels in this community.

The *orange community* (Figure 2(a)) comprises of activities that involve capturing online content and media, and organizing personal knowledge. Typical activities include:

- Various to-do list services as action and trigger channel
- Posting of Web content and URLs including YouTube videos to content management and note taking services
- Google Gmail as action and trigger channel

The *yellow community* (Figure 2(a)) comprises of activities that engage online social networking services and utilize sharing functions for posting content to such services. Activities in this community include:

- Transferring shared content to personal stores
- Broadcast of content or status update to multiple social networking services
- Send a status update on one service about the posting of content on another
- Filter content on one service (e.g. tagged) and pushed the results onto some other service

The *blue community* (Figure 2(b)) comprises of services that orbit around the use of email:

- Email digests
- Event-triggered email notifications (e.g. eBay, Craiglist, reddit)
- Feed as action channel

In Table 3 we identify all the cases where channels of the same service type have been assigned to different communities by the community detection algorithm. For these channels we wish to understand why one channel has an affinity with a particular community while other channels are associated with another.

Service Type	Detected Community				
	Green	*Magenta*	*Orange*	*Yellow*	*Blue*
Email			Google Gmail		Email
Cloud storage		Google Drive		OneDrive, Dropbox	
Phone call	Phone Call	Android Phone Call			
Geolocation	Android Location	iOS Location			
Photos	500px			Instagram, iOS Photos, Flickr	

Table 3: Similar IFTTT channels in different communities

Google Gmail vs Email
The inboxes of both Gmail and other email accounts are the final destinations for content however users in the orange community are using Gmail to relay content to other services (e.g. Picasa/Google+, Kindle, to name a few). This is possible as some online service providers provide an email address for posting content to respective user accounts.

Google Drive vs others
Google Drive channel is used primarily as a general-purpose store while other storage services are primarily used to store content read from social networking services.

Phone Call vs Android Phone Call
Phone Call channel is used primarily as an action channel (i.e. make a phone call) while the Android Phone Call channel is only available as a trigger channel were different types of call events can be used to trigger different actions. The data flow from the latter channel is destined to Google services in the magenta community.

Android Location vs iOS Location
The different affinities for the location-based channels are a consequence of their different use cases. The Android Location channel is strongly paired with both Android Device and Android Notification channels while the iOS Location channel is paired with Google Calendar and Google Drive.

500px vs others
The photographer community site, 500px appears to be misplaced given that all other photo sharing services are assigned to a different community. However when we examine closely the edge weights for recipes involving the 500px channel we discover that the most popular recipe is associated with the Android Device action channel. The particular action interface used allows users to automate a change of the device wallpaper. It is the popularity of the latter function that is responsible for 500px channel’s affinity to the Green community. Moreover users are most likely to set photos from a professional community as their device wallpaper when compared to other photo sharing services which have a stronger affinity with social networking services.

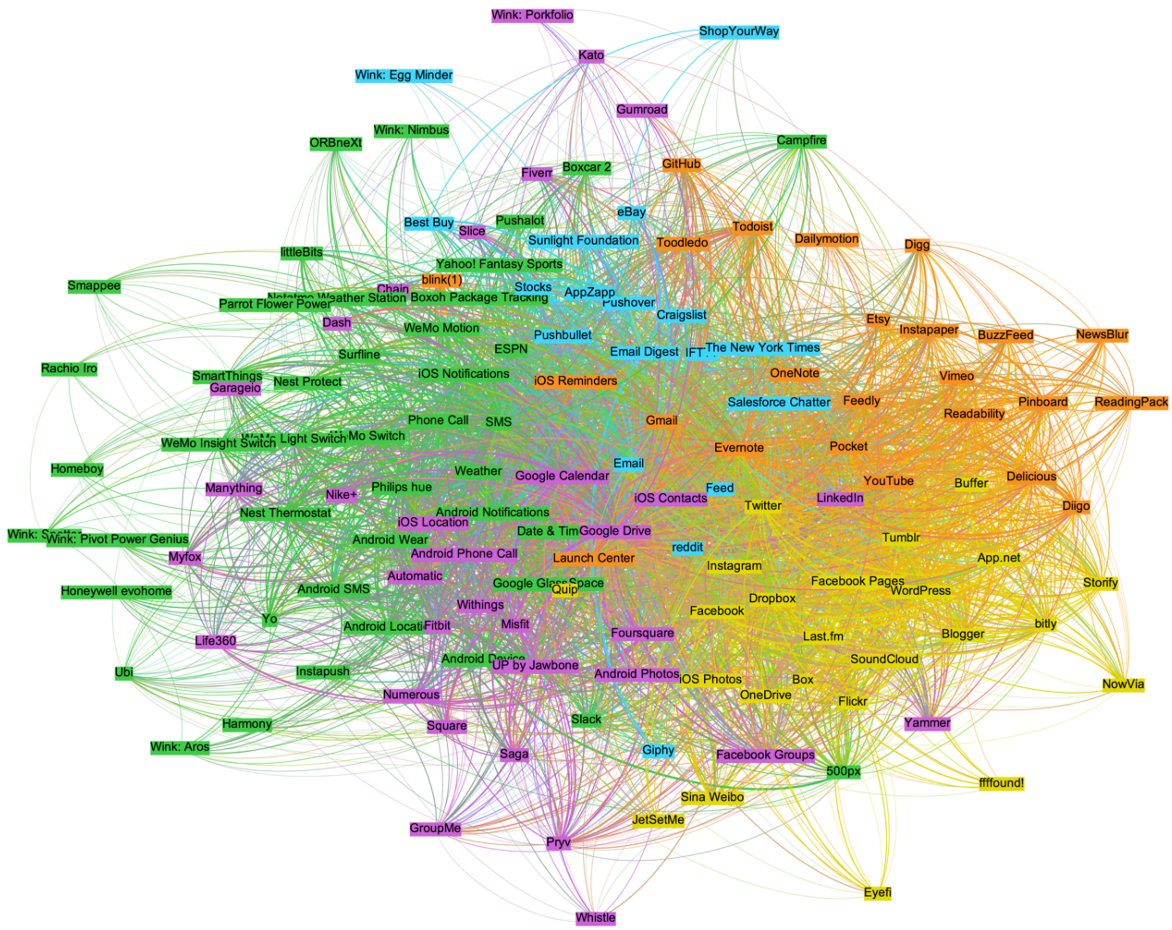


Figure 1: Communities in IFTTT channel graph

Community Attributes	*Highest weighted in-degree*	*Highest weighted out-degree*	*Most popular recipes*
Green community “Internet of Things” 45 nodes (31.25%) 381 edges (9.98%)	Android Notifications (308231) Android Device (288986) SMS (262757) iOS Notifications (192839) Philips hue (127340)	Weather (580894) Date & Time (257490) Android Location (234828) SMS (93024) Space (57611)	Android Location – Android Device (133014) Weather – Android Location (122137) Weather – iOS Notifications (109945) Weather – SMS (72949) Android Location – Android Notifications (66166)
Magenta community “Quantified self and assets” 33 nodes (22.92%) 165 edges (4.32%)	Google Drive (590687) Google Calendar (214414) UP by Jawbone (39338) LinkedIn (31078) Numerous (4224)	iOS Contacts (174614) iOS Location (100688) Foursquare (78997) Google Calendar (69013) Android Phone Call (56427)	iOS Contacts – Google Drive (130773) Foursquare – Google Calendar (38313) Android Phone – Google Drive (26645) iOS Location – Google Drive (20089) iOS Location – Google Calendar (18686)
Orange community “Knowledge management” 24 nodes (16.67%) 218 edges (5.71%)	Evernote (438571) Pocket (200984) Gmail (144217) iOS Reminders (50055) Feedly (30752)	Gmail (292223) Pocket (122708) YouTube (99662) Feedly (88671) iOS Reminders (64012)	Gmail – Evernote (74280) Pocket – Evernote (70913) YouTube – Pocket (44706) Feedly – Pocket (38158) Gmail – iOS Reminders (34174)
Yellow community “Social, music” 24 nodes (16.67%) 282 edges (7.39%)	Twitter (56262) Dropbox (457477) Facebook (157326) iOS Photos (100438) Tumblr (90755)	Instagram (444749) Facebook (433760) iOS Photos (281293) Twitter (143025) Dropbox (67090)	Facebook – Twitter (154318) Facebook – Dropbox (125901) Instagram – Twitter (109721) Instagram – Dropbox (103080) iOS Photos – Dropbox (55722)
Blue community “Email and feeds” 18 nodes (12.5%), 56 edges (1.47%)	Email (872747) Email Digest (58333) Pushbullet (55291) Pushover (45341) reddit (2527)	Feed (912676) IFTTT (158286) Craigslist (85650) Email (77796) reddit (58634)	Feed – Email (370159) IFTTT – Email (127547) Craigslist – Email (5516) reddit – Email (21374) Feed – Pushbullet (16287)

Table 4: Most popular trigger/action channels and recipes in each community

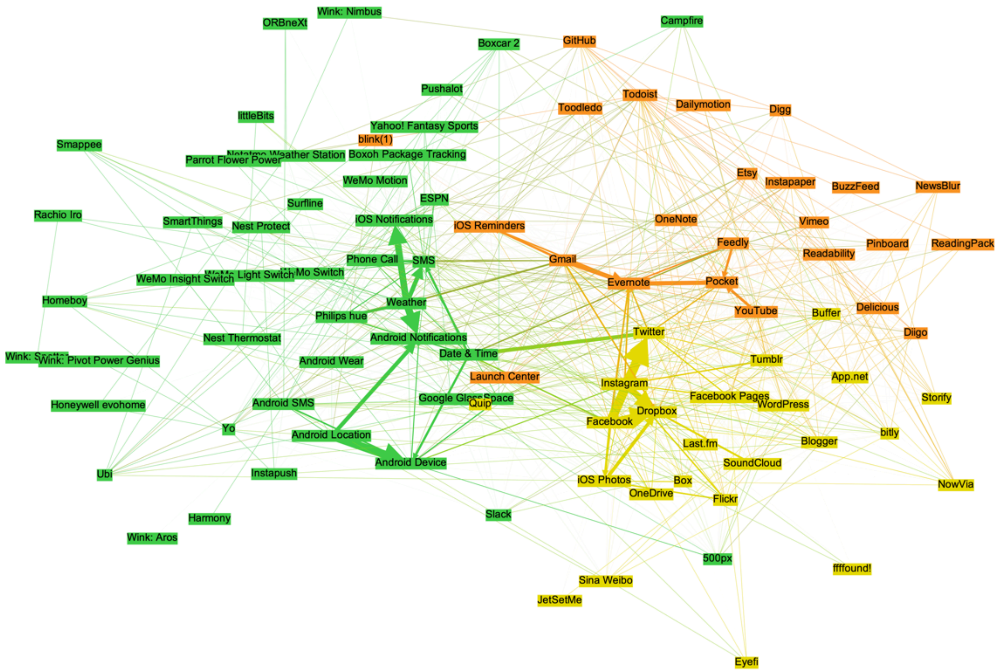
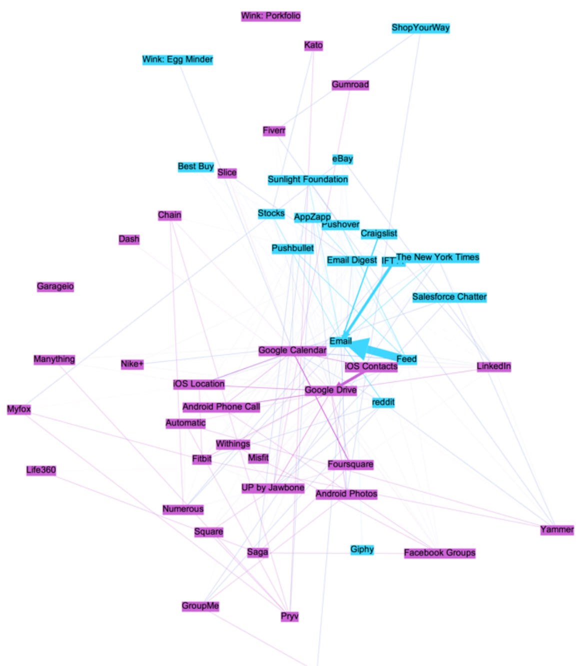
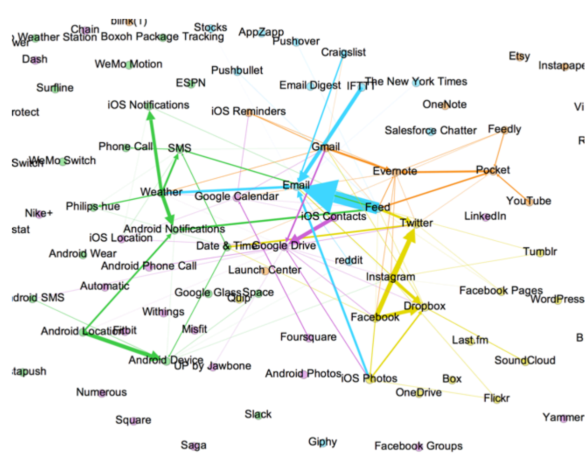


Figure 2(a): Green, orange and yellow communities





Figures 2(b): Magenta and blue communities; 2(c) Core of the channel graph

C. Average Path Length, Mean Channel Engagement and Betweenness Centrality

Using Gephi we determine that the diameter of G is 3 nodes and the average path length is ~1.67 hops. This is corroborated by the observation that popular IFTTT recipes are skewed towards an *inner core* of channels. These recipes represent the most popular interactions with smartphones and online services (Section VI). Moreover recipes tend to combine popular multi-use channels with a specialized channel. The compactness of G is a consequence of a significant proportion of users engaged with at most two channels. From user data we determine that the *mean channel engagement* is ~2 and the mean recipes per channel is 0.73. A plot of the number of users versus channel count is shown in Figure 3 – most users have a channel engagement of 5 and below and a small minority of users have a channel engagement of >20. Although this plot suggests a general tendency of users to engage with a few channels there also exists stark differences in usage patterns for the most active users of IFTTT: for the top two users sharing the most recipes – one user has created 568 recipes over 3 channels while the other has created 489 recipes over 93 services.

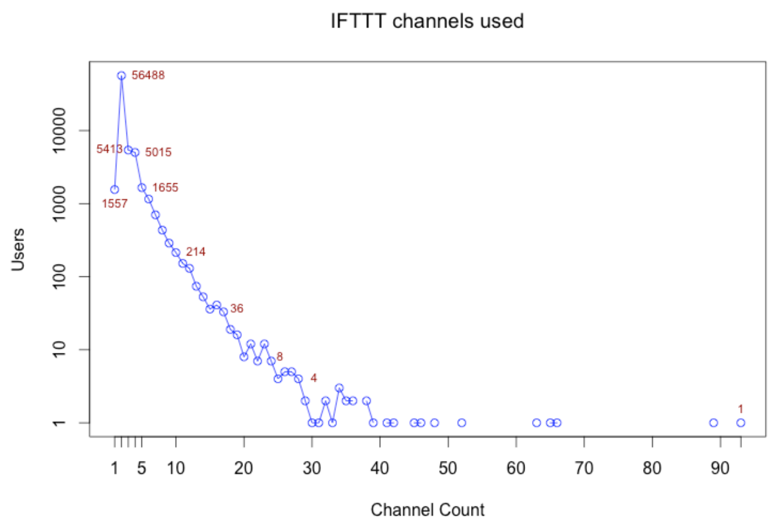


Figure 3: Plot of channel engagement

We calculate the betweenness centrality metric [9], which is an indicator of a channel's centrality in G (Figure 4(a)). We interpret a channel with high betweenness centrality as a major hub where a significant proportion of IFTTT recipes converge. In other words, the channel is used in popular recipes as both a trigger and action function. Note that this metric assumes that there exists a flow of data between channels along a particular path and such data flows follow the shortest paths. It is evident from the public descriptions of recipes that a high proportion of data flows for a particular recipe do not join up with flows from other recipes, however it is not obvious whether some users are using one recipe to trigger another. Nevertheless this metric does identify candidates for potential connection points where multi-channel spanning data flows might transit.

We can classify channels according to whether they are used as a trigger and/or action function:

- *Sources:* 50 channels (34.73%) have zero weighted in-degree and betweenness centrality
- *Sinks*: 29 channels (20.14%) have zero weighted out-degree and betweenness centrality
- *Hubs*: 65 channels (44.14%) have non-zero in-degree and out-degree

Sources and sinks represent termination points in G while hubs are channels that are can be sources and sinks of data. Sources are channels solely used in a read-only capacity and represent generators of content and media while sinks are channels used in a write-only capacity and tend to represent the final destination of content or status notification. Figure 4(b) depicts the SMS hub channel (left) with the highest betweenness value of 969.019; this is contrasted with the Feed source channel (right) with a betweenness value of 0 and the highest weighted out-degree of 912676. Figure 4(c) shows the contrasting uses of two online storage services: Google Drive sink channel has a betweenness value of 0 and the second highest in-degree of 590687 while the Dropbox hub channel is used as a sink (457477) primarily and as a source (67090).

Channel	*Betweenness Centrality*
SMS	969.019
Twitter	815.269
Email	685.396
Gmail	681.436
Google Calendar	656.191
Facebook	548.09
Dropbox	341.437
Tumblr	210.226
UP by Jawbone	149.268
WordPress	122.328

Table 5: Top 10 channels with the highest betweenness centrality values

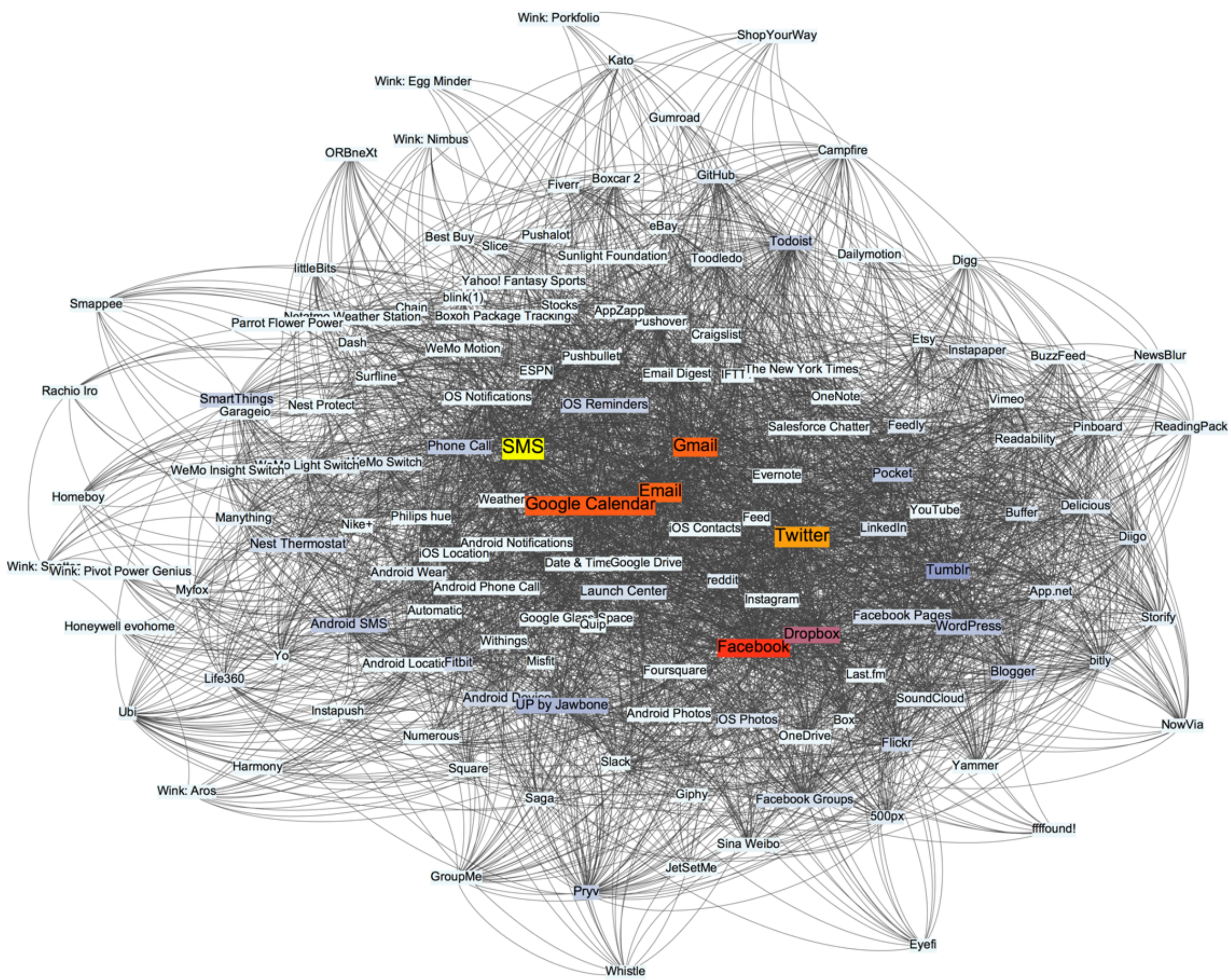


Figure 4(a): Betweenness centrality graph: large to small values mapped using heat map color palette (yellow – orange – red – gray – white) and nodes are sized proportionately.

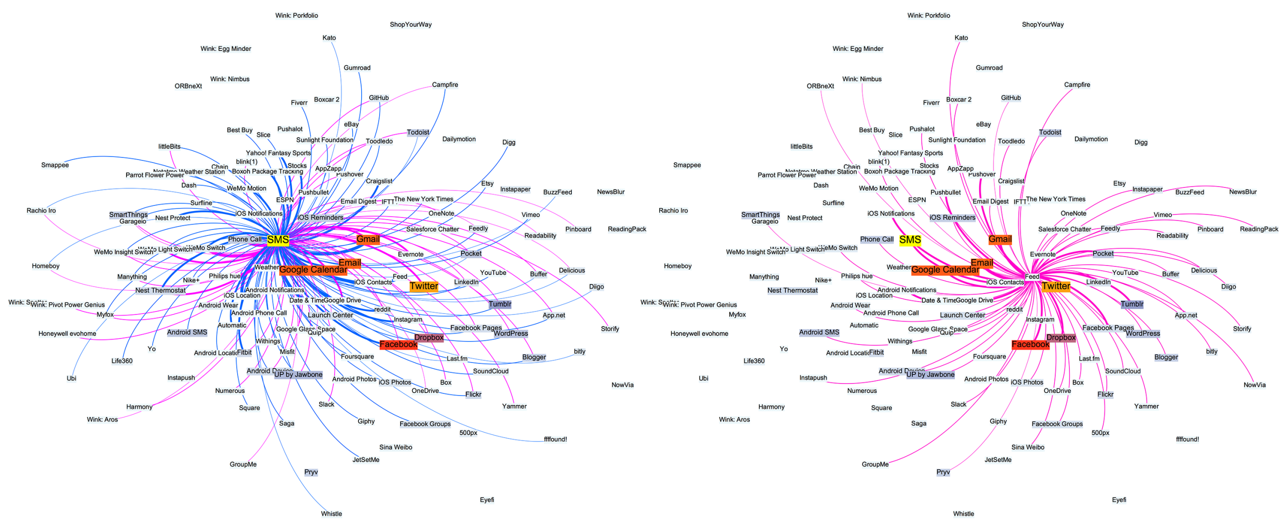


Figure 4(b): Betweenness centrality graph: *SMS* (right) and *Feed* (left) channel –
incoming edges in blue and outgoing edges in magenta, thicker lines indicate higher edge weights

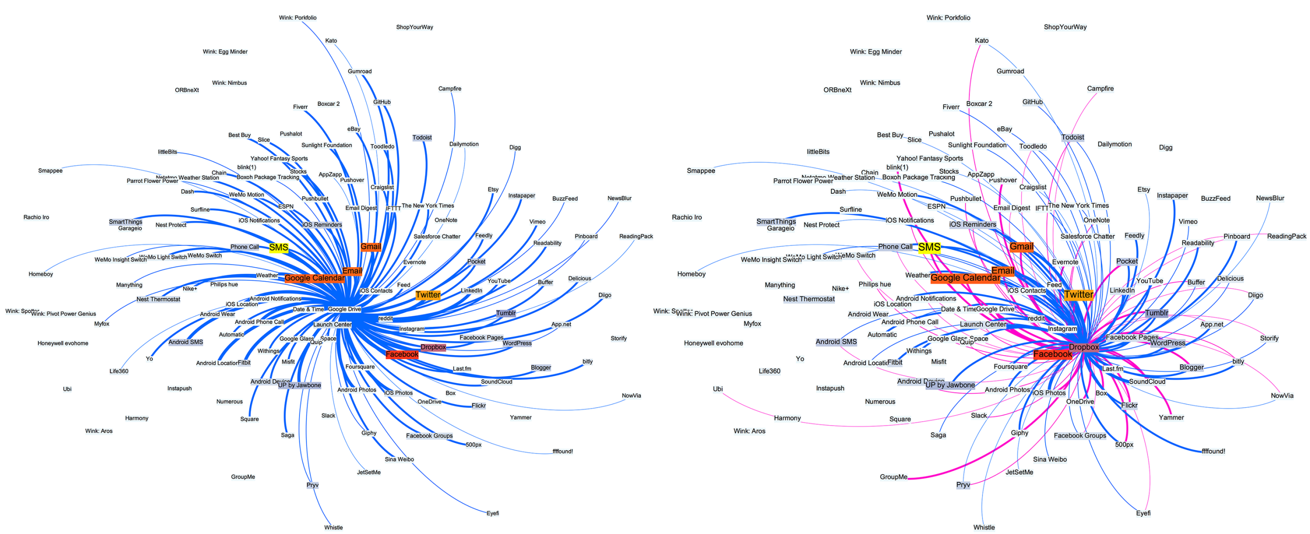


Figure 4(c): Betweenness centrality graph: *Google Drive* (right) and *Dropbox* (left) channel

V. INNER CORE OF CHANNELS VISUALIZATION

Using Gephi with the graph layout settings of proceeding sections, we also make the following modifications in order to emphasize the *inner core* of IFTTT channels:

- *heat map* color scheme on dark background (color nodes with palette transition yellow – orange – red – gray – white for nodes ranging from higher to lower weighted degrees)
- node size is based on weighted degree
- filter out directed edges (s, t) with edge weights $k_{st} < 100$
- edge thickness is based on weights k_{st} (rescaled in Preview mode)

We render 4 graphs where the inner core comprises of nodes with the highest non-weighted out-degree (Figure 5(a)), non-weighted in-degree (Figure 5(b)), weighted out-degree (Figure 6(a)) and weighted in-degree (Figure 6(b)). We filtered directed edges to show channels acting as either sources or sinks (Section IV-C) and we also cull low-use channels for clarity of the visualization.

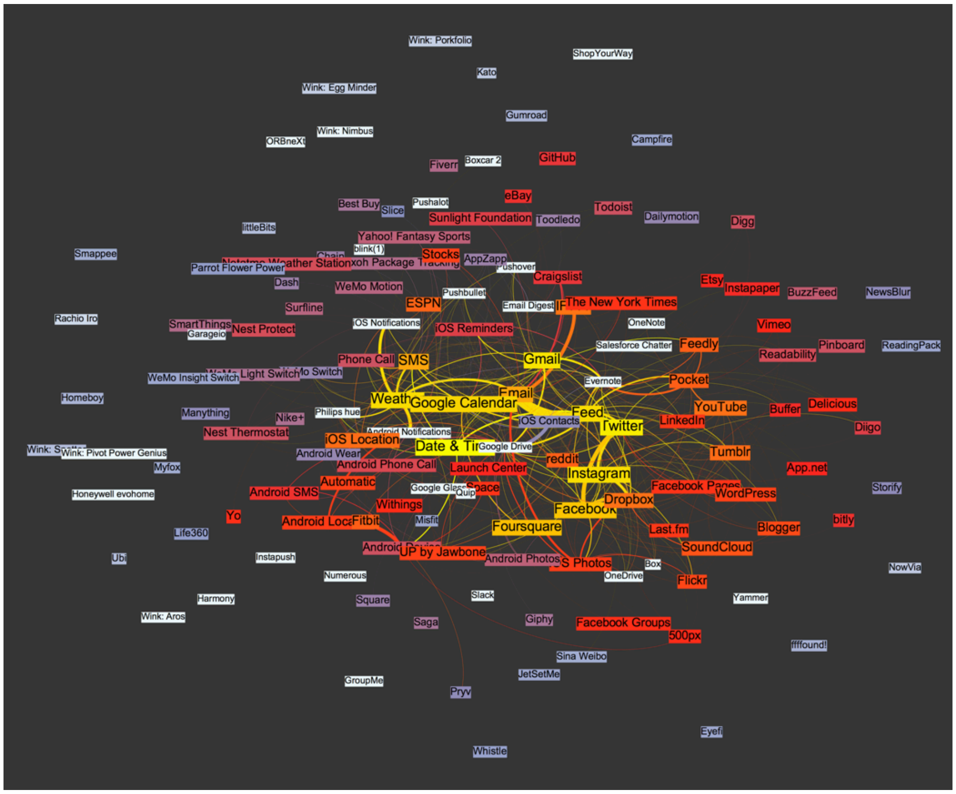


Figure 5(a): Inner core of channels $k_{st} \geq 100$ (non-weighted out-degree)

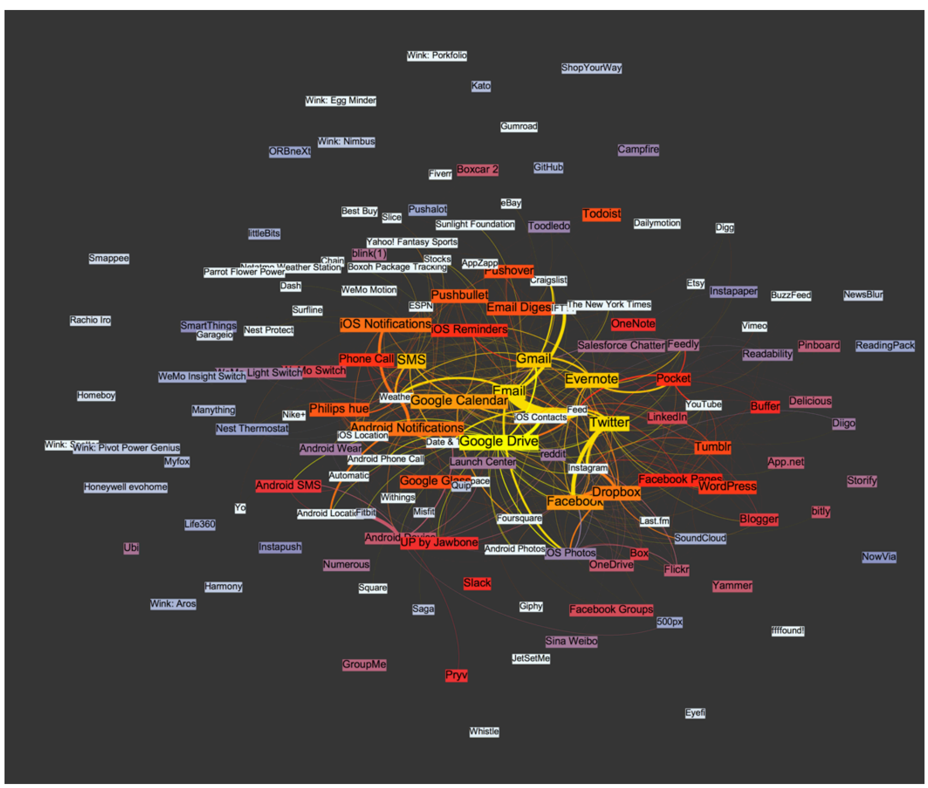


Figure 5(b): Inner core of channels $k_{st} \geq 100$ (non-weighted in-degree)

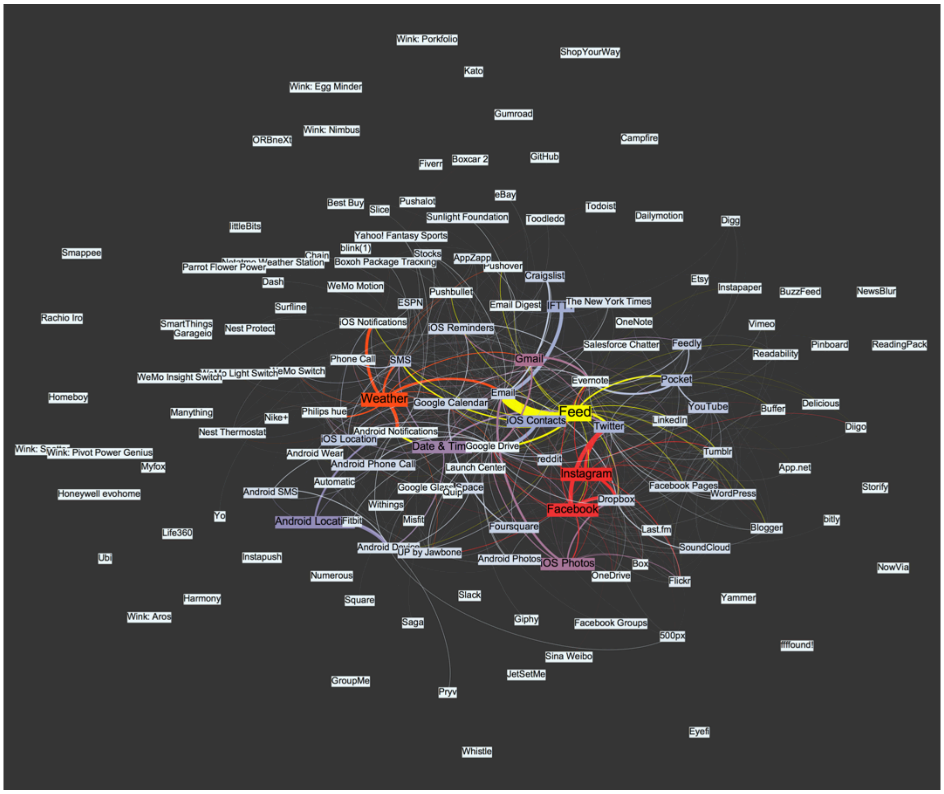


Figure 6(a): Inner core of channels $k_{st} \geq 100$ (weighted out-degree)

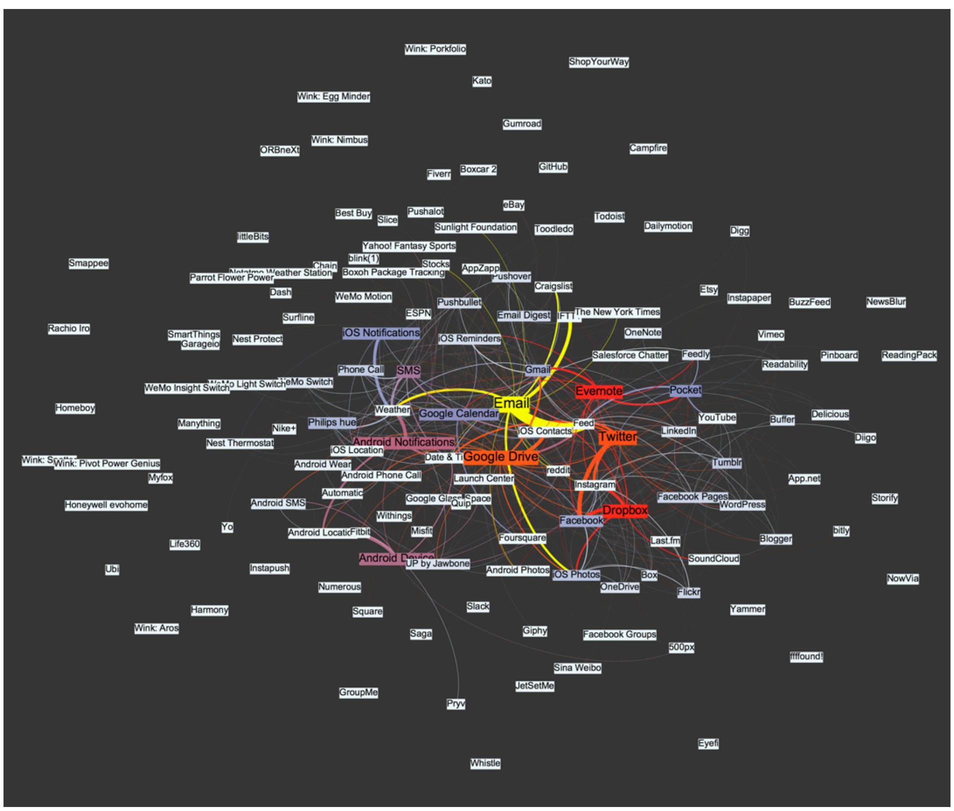


Figure 6(b): Inner core of channels $k_{st} \geq 100$ (weighted in-degree)

The visualizations based on non-weighted degree values (Figure 5) reveals the *breadth* of recipes and is indicative of the possible use cases for channels – in terms of usage in a recipe[5], Date & Time and Google Drive channels represent the most popular action and trigger channels respectively. The weighted degree values measures the popularity of channels used as either a trigger and/or an action. The weighted degree value is based on the weights k_{st} of incoming or outgoing directed edges. It is evident from visualizations of Figure 6 that a small percentage of channels appear in a majority of user recipes – in terms of user add count, Email and Feed channels represent the most popular action and trigger channels respectively. These visualizations indicate that despite a wide spread of available recipes involving a certain channel, a few recipes dominate and these collectively represent the most popular recipes in the IFTTT service.

VI. CONCLUSIONS

We have revealed distinctive usage patterns for IFTTT channels based on the channel graph of Section IV using graph metrics and community detection and betweenness centrality algorithms. A detailed examination of the membership of each community indicates patterns of use that revolve around certain services and each community correlate with characteristic hubs of activity between users and their devices. As the IFTTT tasking service grow in terms of the number of channels and user popularity these detected communities might reinforce themselves or perhaps there might exist new services or functionality that might reconfigure the memberships of each community.

The analysis has revealed that most recipes tend to be isolated pairwise combinations of channels although we do not have information on whether some users are transforming data flows with more than one recipe. The concatenation of recipes into longer chains can be crudely compared to a construction of a Rude Goldberg machine [10], which suggests significant cognitive load in arranging recipes that join up in longer chains.

The analysis only captures user and machine activity as indirectly measured by the recipes users have created. Furthermore we do not have information on how often recipes are used and how recipes are scheduled. Our data does not include activity outside of IFTTT's network. Nevertheless this restricted study provides a useful template on how to organize data collection and analysis of similar tasking services and the results of the study may reflect prevailing usage patterns not brokered by IFTTT – the number of ways in which a service can be used is restricted by API functionality and other limitations on usage and these restrictions are agnostic of the tasking network.

VII. REFERENCES

[1] About IFTTT (https://ifttt.com/wtf)

[2] IFTTT recipes (https://ifttt.com/recipes)

[3] IFTTT channels (https://ifttt.com/channels)

[4] Gephi, Open Graph Viz Platform (https://gephi.github.io)

[5] R Project (http://www.r-project.org)

[6] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, Etienne Lefebvre, *Fast unfolding of communities in large networks*, in Journal of Statistical Mechanics: Theory and Experiment 2008 (10), P1000

[7] Wikipedia article: https://en.wikipedia.org/wiki/Internet_of_Things

[8] Wikipedia article: https://en.wikipedia.org/wiki/Quantified_Self

[9] Ulrik Brandes, *A Faster Algorithm for Betweenness Centrality*, in Journal of Mathematical Sociology 25(2):163-177, (2001)

[10] Wikipedia article: https://en.wikipedia.org/wiki/Rube_Goldberg_machine

[5] Note that this ignores unique recipes incident on a directed edge