

Web and Social Media Analytics

Social media (core concepts and data flow)

Dr Philip Worrall

School of Computer Science and Engineering
115 New Cavendish Street
University of Westminster
London, W1W 6UW
worrallph@westminster.ac.uk

LW7

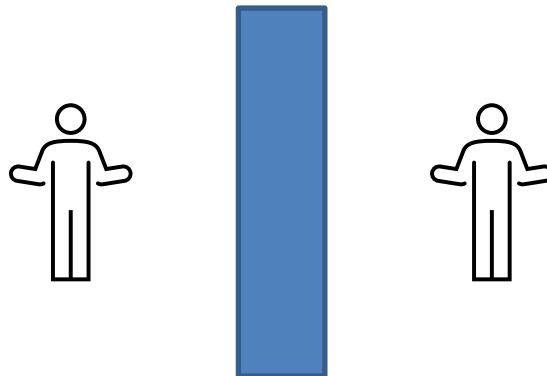
Week 7

Plan for today

- ❖ Core concepts
- ❖ The rise of social media
- ❖ Application areas
- ❖ Data flow

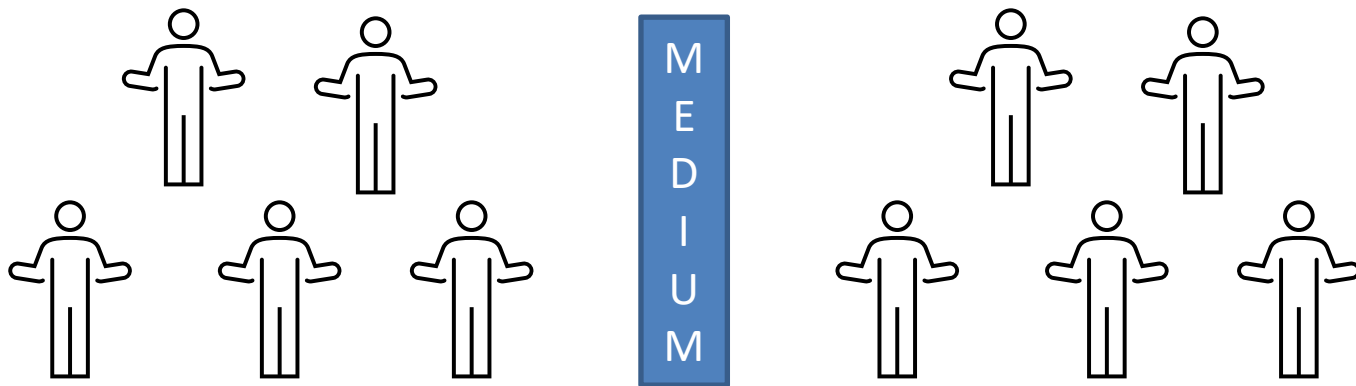
Social media

- Media is the plural of **medium**
- Medium
 - A middle state or condition
 - Something intermediate in nature
 - Closely related to the word *median* → the middle



Social media

- Social meaning **companionship** or **living** with others in a community, rather than in isolation
- “Given the importance of social interactions for humans, it is not surprising that most psychiatric disorders involve some disruption of normal social behavior” (Young, 2008)



Social media

- Key characteristics of social media
 - I. Enables broadcasting of ideas or thoughts
 - II. Encourages or facilitates collaboration
 - III. Aids formation of communities or relationships
- Social media exists in a variety of different forms
- What forms of offline social media can you think of?

Online social media

This is the **contemporary** form of social media that many of use now use.

In this case the medium is the online platform, which itself is hosted on the web.

A relatively new phenomenon, having not existing at the beginning of the web.

In part, this is explained by its reliance on the use of **enabling technologies**.

Social media technologies

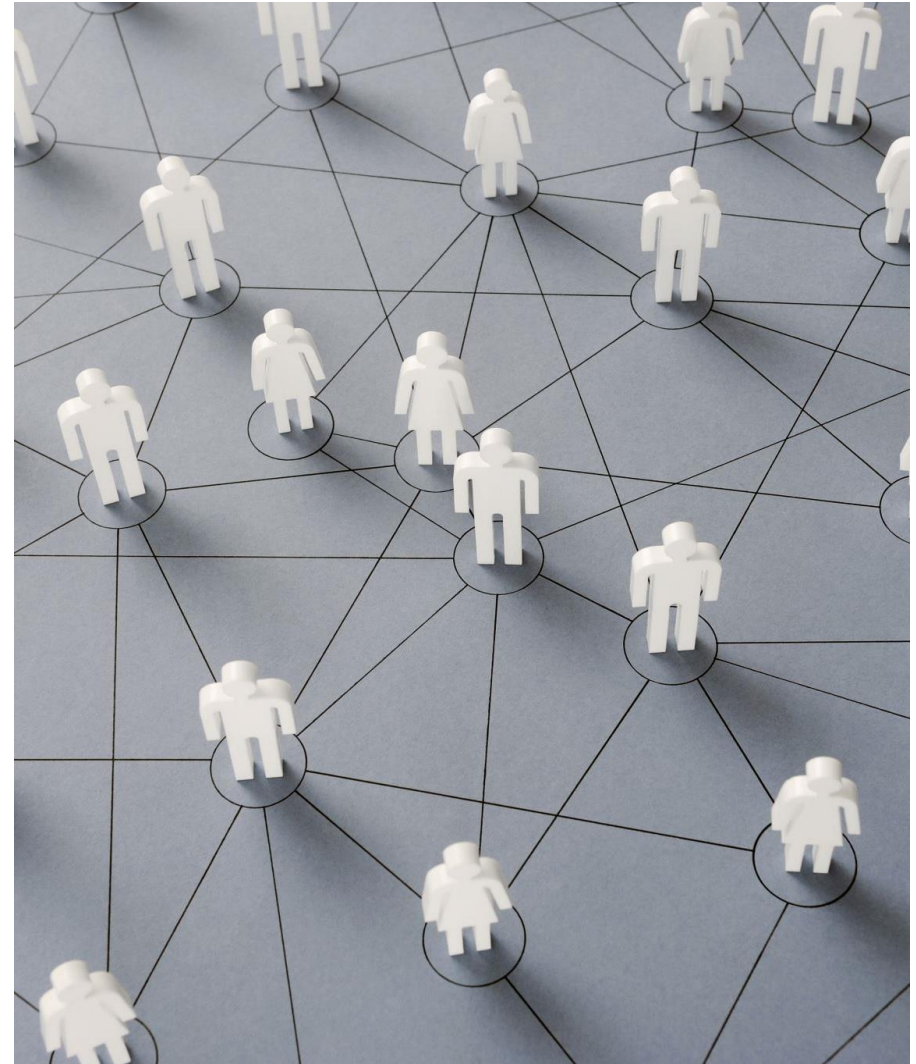
- It could be argued that a key enabler of SM has been the adoption of **fast internet** (broadband) and **web programming**
- Adoption has therefore closely followed the rollout of 3G and xDSL around the globe.
- Key SM features:
 - Ability to comment and submit reviews/ratings
 - File sharing (e.g., music, videos or photos)
 - User-edited content (wiki, profiles, blogs)

Social media networking

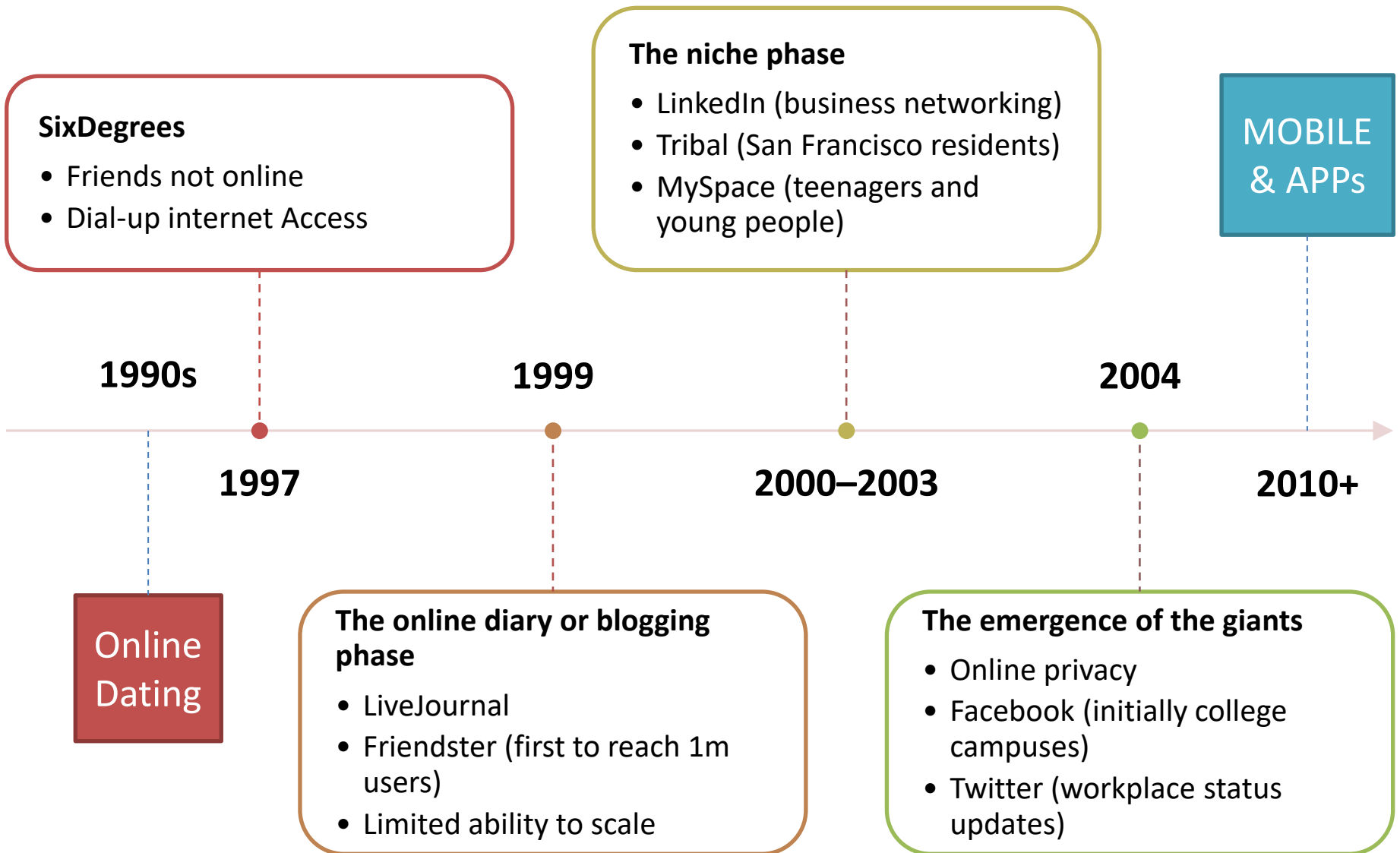
A **special class** of social media sites

Enables individuals to (1) construct a public or semi-public profile within a bounded system, (2) specify names of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system.

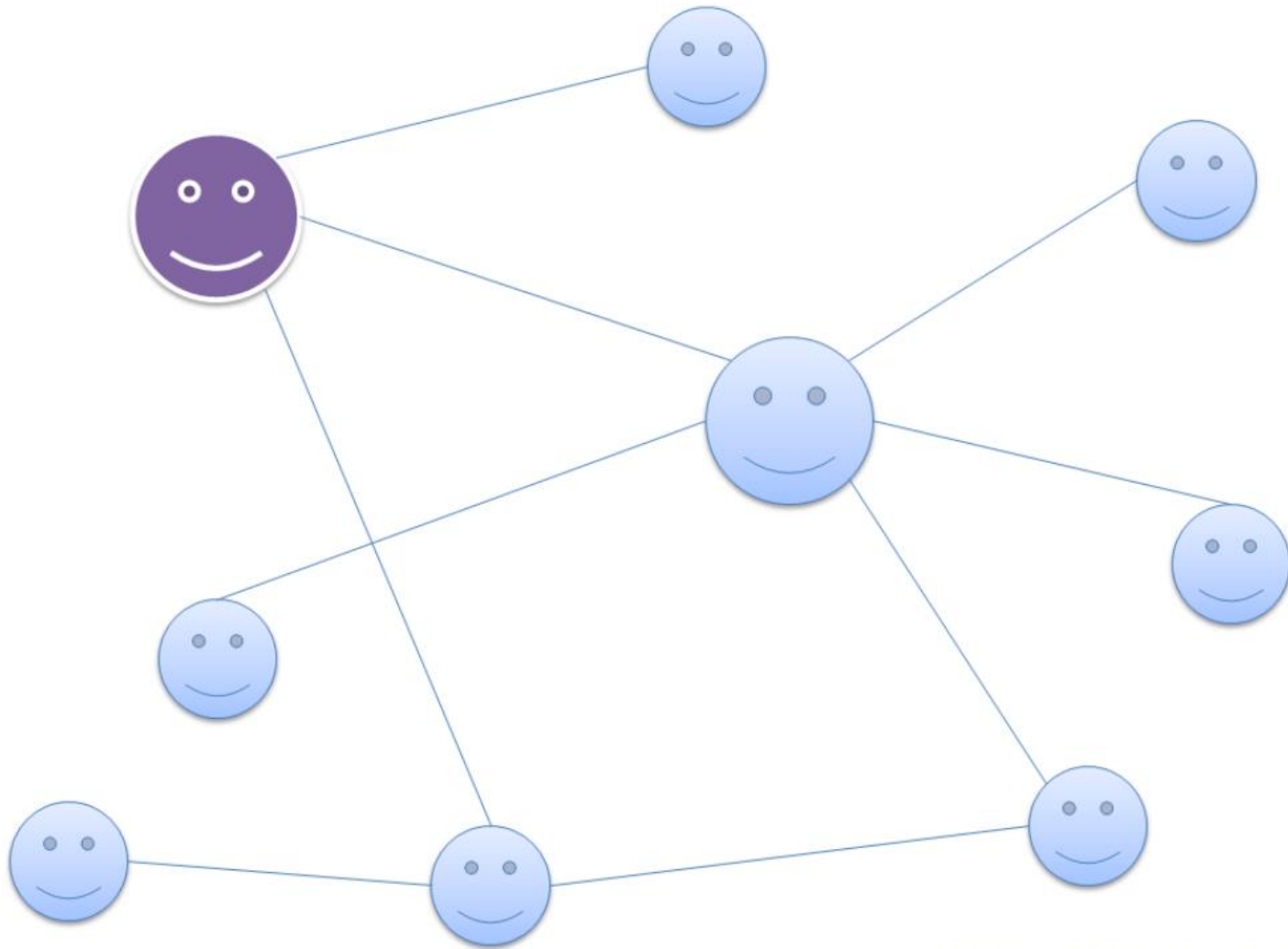
The nature and nomenclature of these connections may vary from site to site.



Social media has evolved in several waves...



SixDegrees



Early Facebook



The screenshot shows the early Facebook website. At the top left is a pixelated profile picture of Mark Zuckerberg. To its right is the text "[thefacebook]" in a large, blue, monospace-style font, with links for "login", "register", and "about" below it. The main content area has a blue header bar that says "Welcome to Thefacebook!". Below this is a large heading "[Welcome to Thefacebook]". The text describes the site as an online directory for colleges, specifically mentioning Harvard University. It lists four uses: searching for people at your school, finding out who is in your classes, looking up friends' friends, and seeing a social network visualization. It then prompts users to register or log in, with corresponding buttons. At the bottom, there are links for "about", "contact", "faq", "terms", and "privacy", followed by "a Mark Zuckerberg production" and "Thefacebook © 2004". On the left side, there is a registration form with fields for "Email:" and "Password:", and buttons for "register" and "login".

[thefacebook]
login register about

Welcome to Thefacebook!

[Welcome to Thefacebook]

Thefacebook is an online directory that connects people through social networks at colleges.
We have opened up Thefacebook for popular consumption at **Harvard University**.

You can use Thefacebook to:

- Search for people at your school
- Find out who are in your classes
- Look up your friends' friends
- See a visualization of your social network

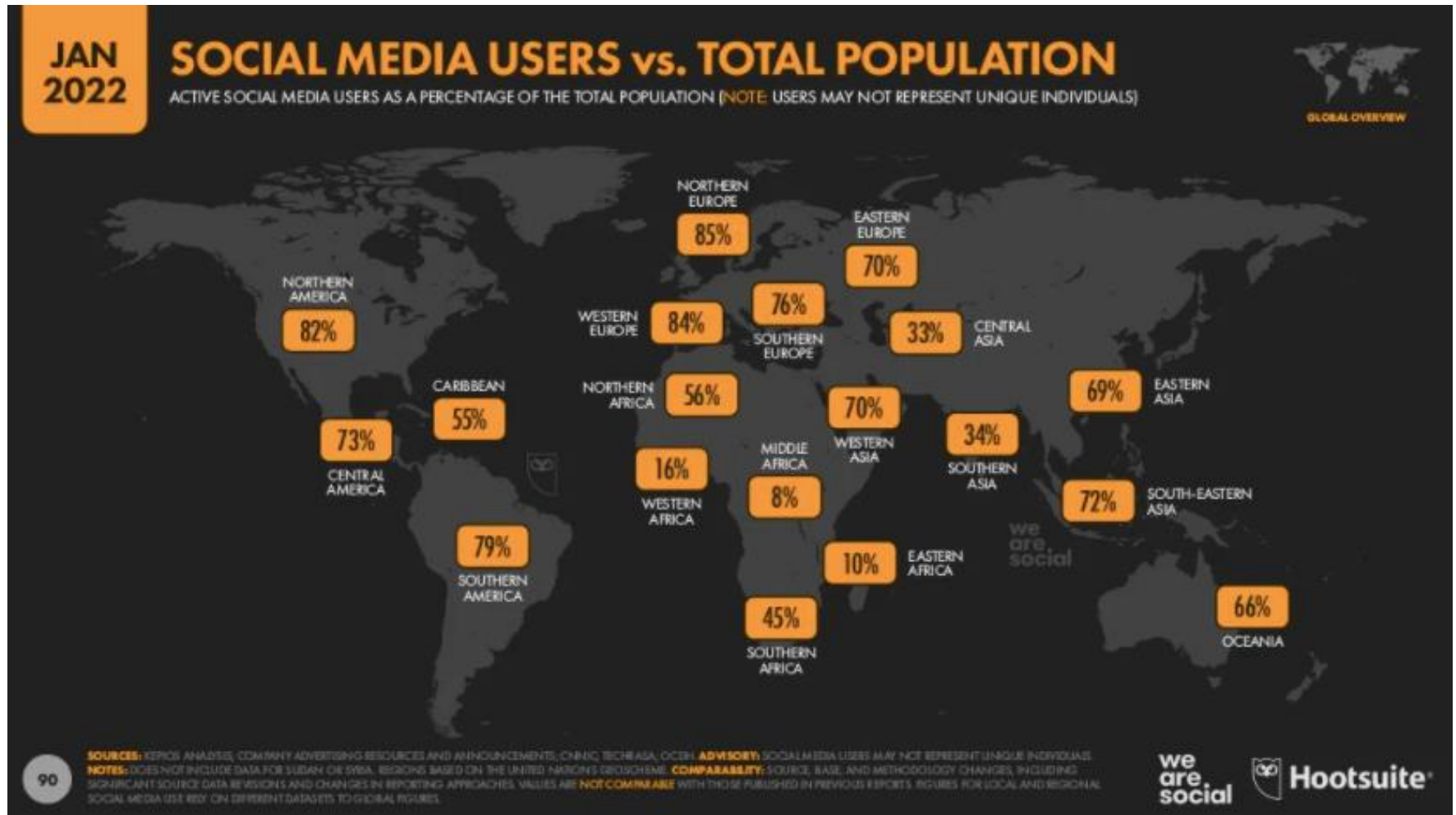
To get started, click below to register. If you have already registered, you can log in.

Register Login

about contact faq terms privacy
a Mark Zuckerberg production
Thefacebook © 2004

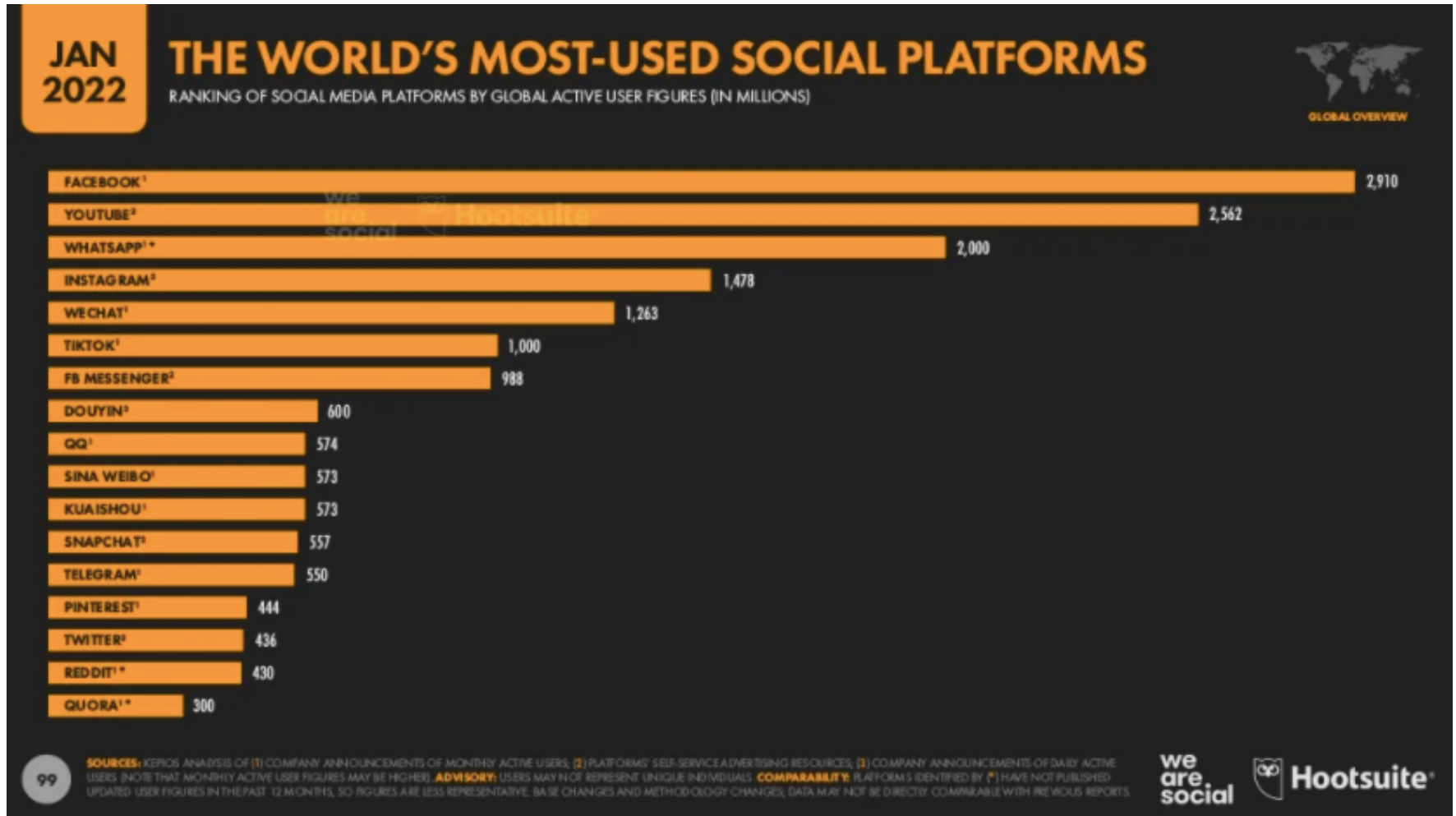
Email:
Password:
register login

Social media adoption



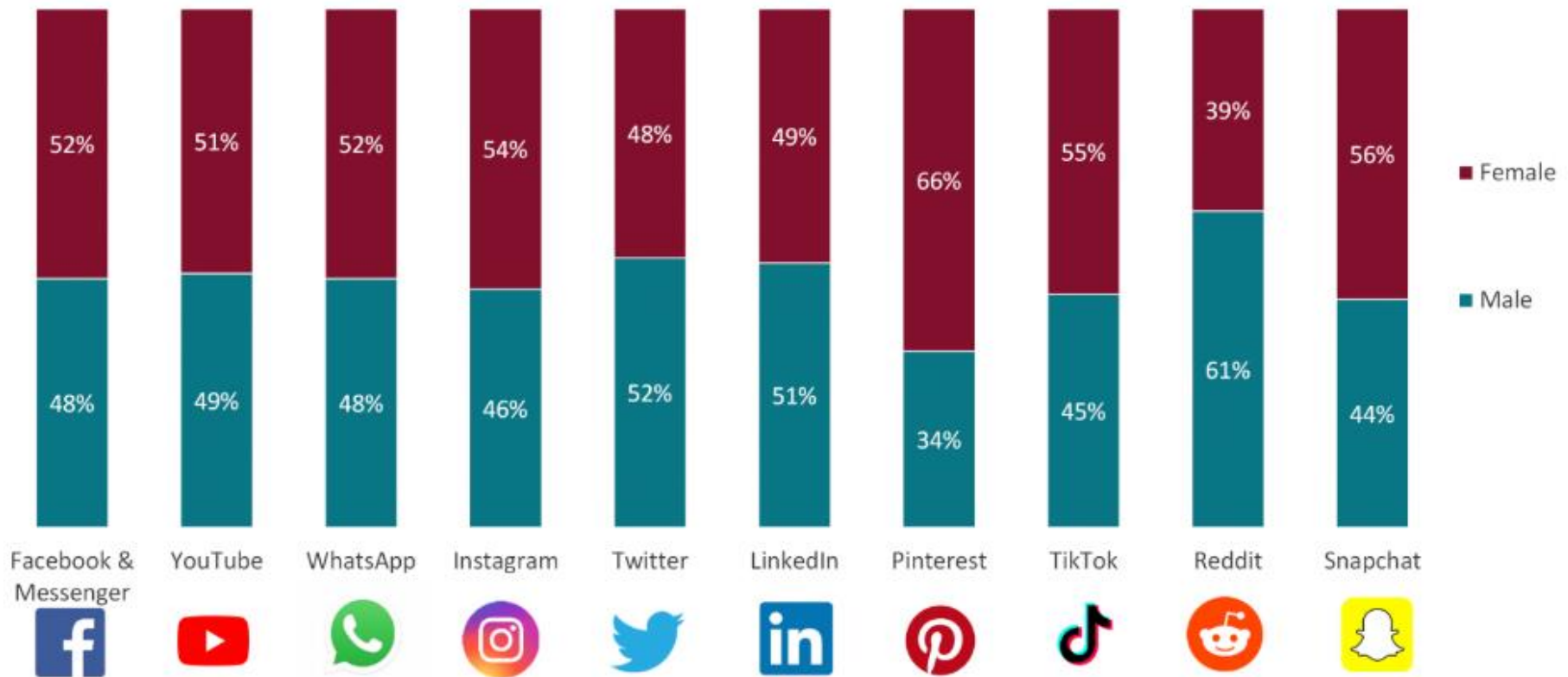
(SmartInsights, 2022)

Social media platforms by active users



(SmartInsights, 2022)

Adoption across social media sites

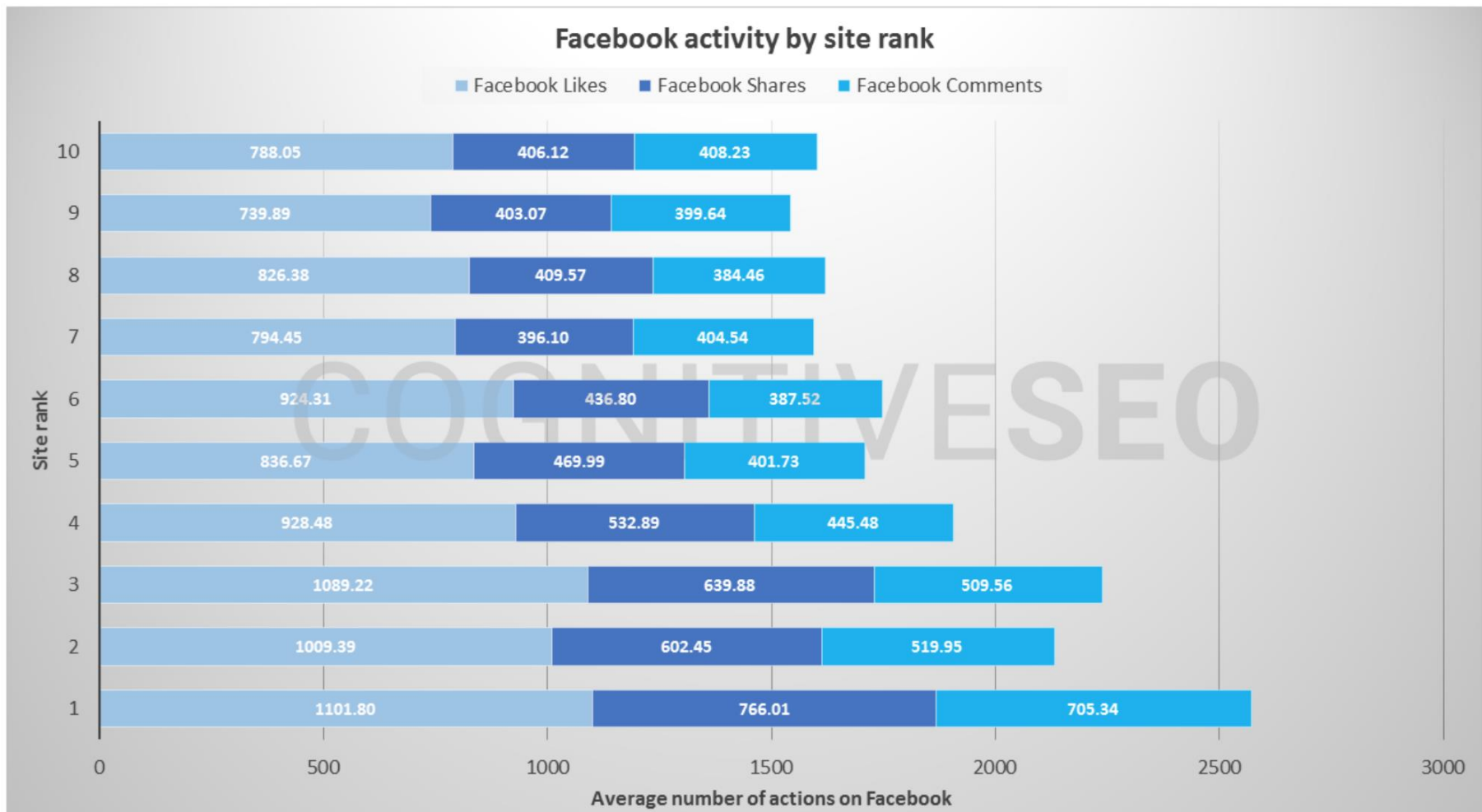


(Ofcom, 2022)

Business Case for Social Media

- Marketing
 - New product launches
 - SMO
- Social search
- Customer service
- Recruitment
- eAlerts
- Crowd funding/eActivism

Social signals and SMO



(Cognitive SEO, 2016)

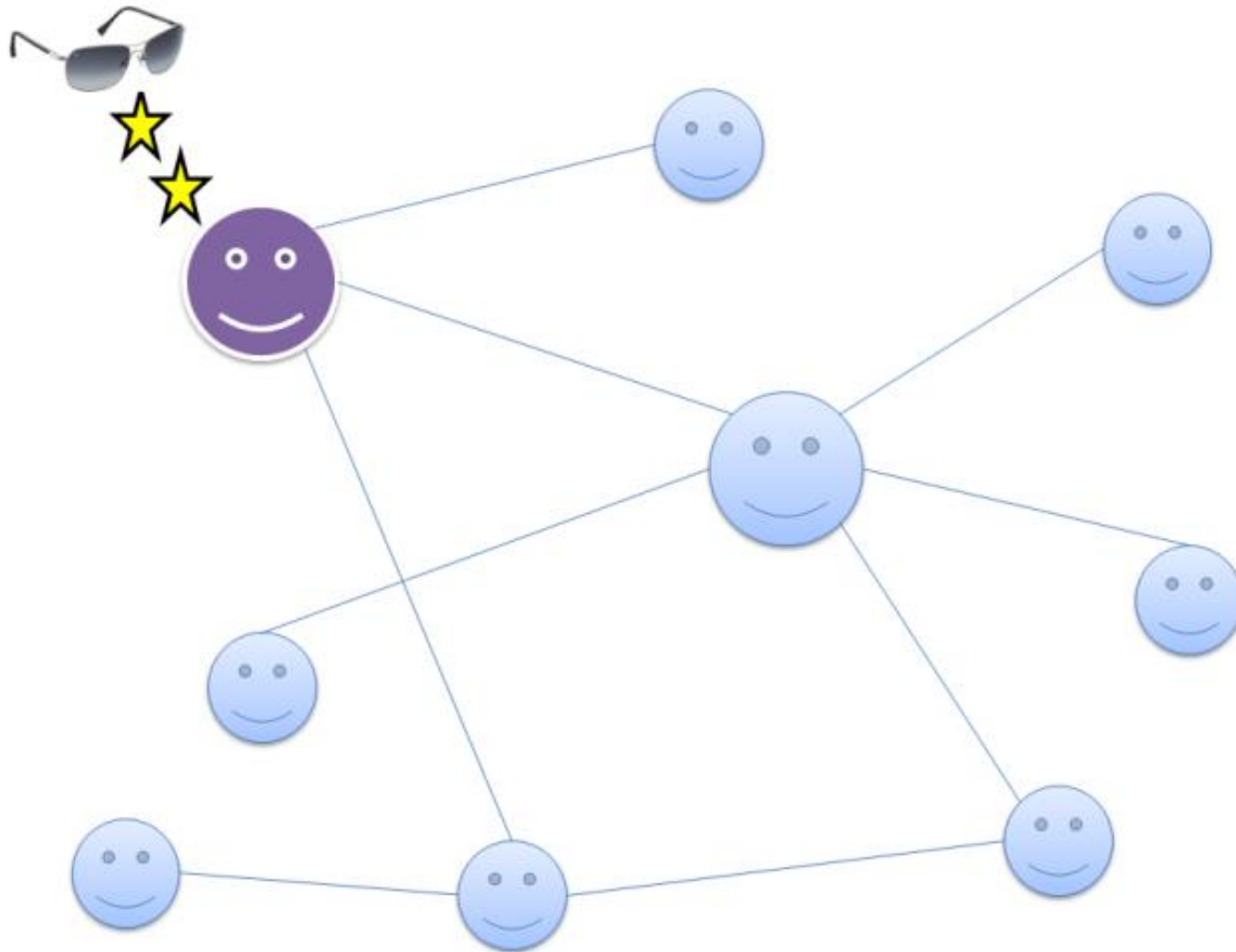
Studying social media

Research Fields

In addition, social media has been used to study a variety of different phenomenon;

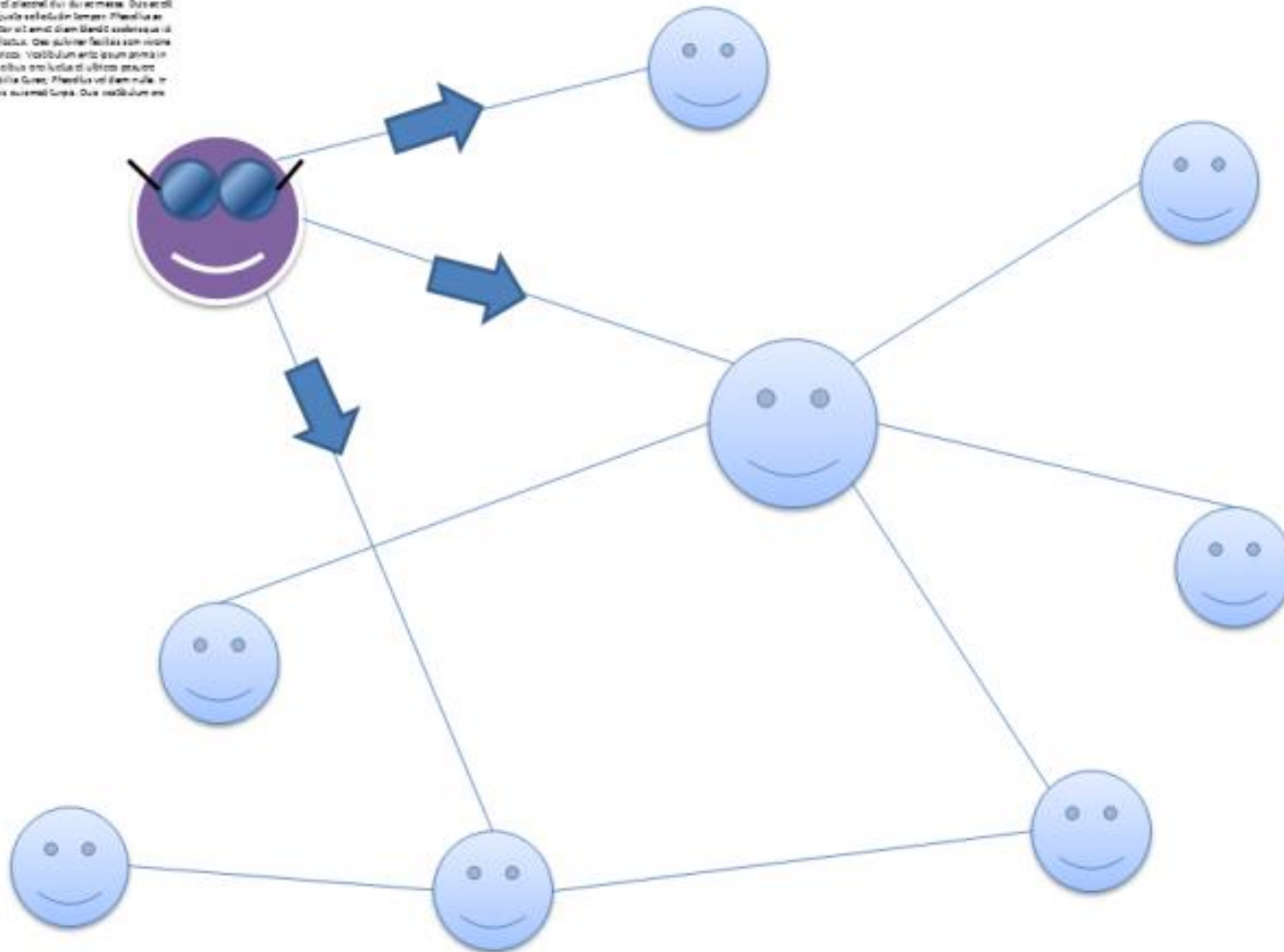
- Formation and Maintenance of Friendships (Psychology)
- Distribution and Communication of Information (MAD)
- Development of Distributed Networks (CS)
- Promotional Space (Business)
- Consumer Opinion and Debate (Intelligence and Analytics)

Case1: SMN

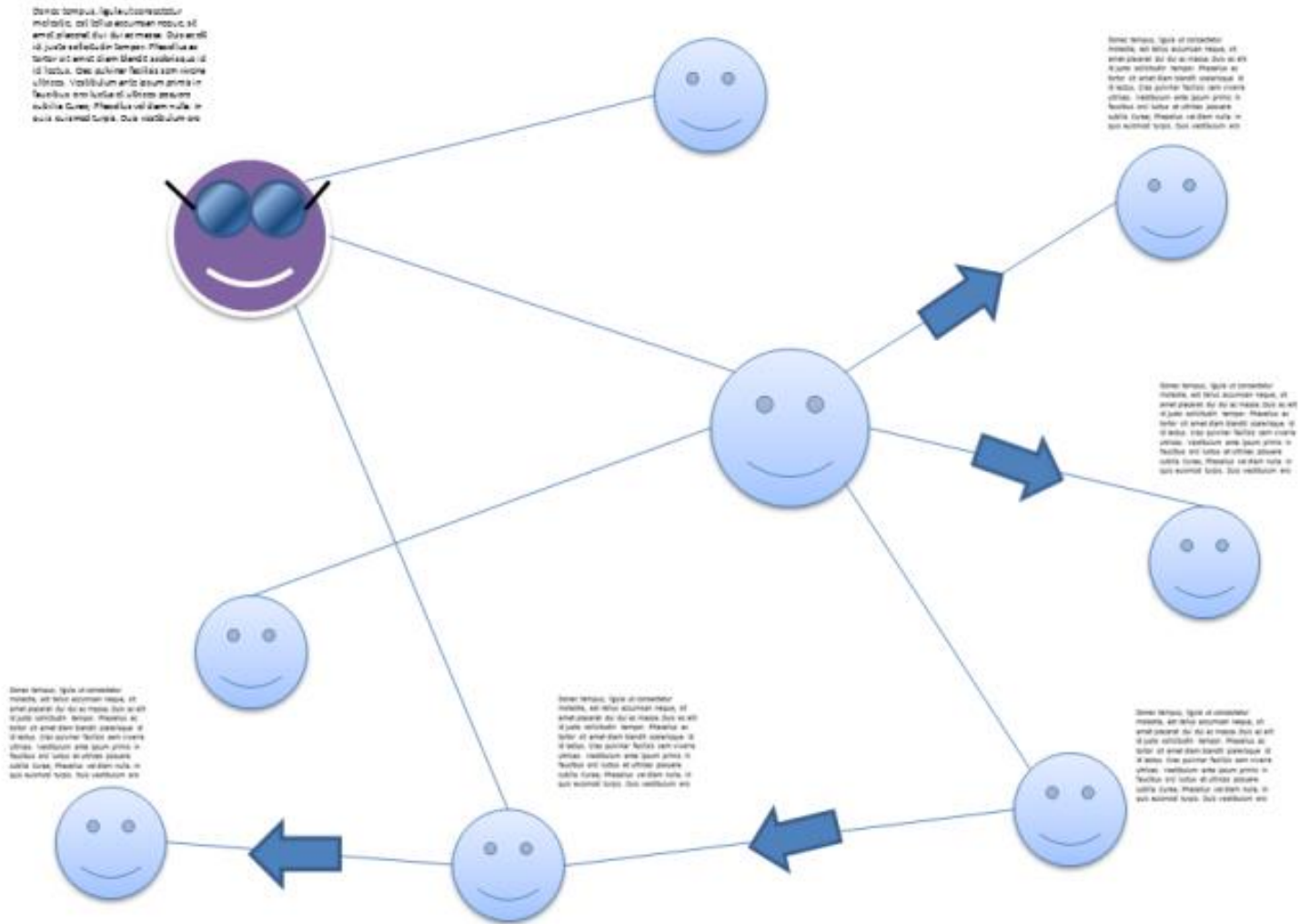


Case1: SMN

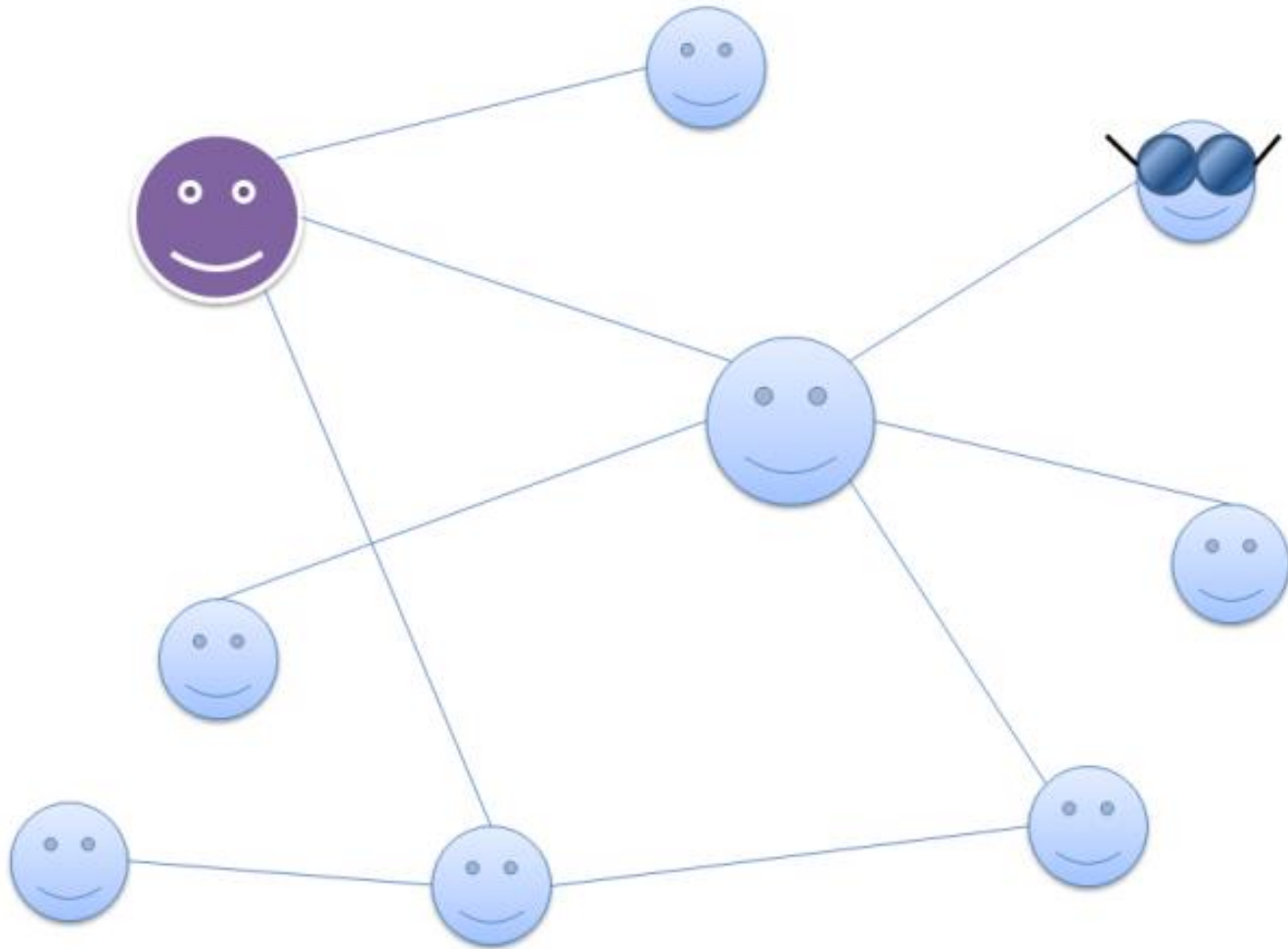
Dennis Lemus, Igulubonnetur
molitio, ut lottuacurman ipius ut
amit dlatent dui dui amesse. Quo anelli
ut pulte velletudin tempen. Pheolus an
tortor ut amot dlam blandi coloris qu ut
id factus. Quo sulcorer facilia som noine
ultrata. Vocabulum ante quam amia in
teutibus pro lullus ut ultione amato
mullia Gure; Pheolus vel dlam nulla. In
eura euraemul Gure. Quo vocabulum an



Case1: SMN

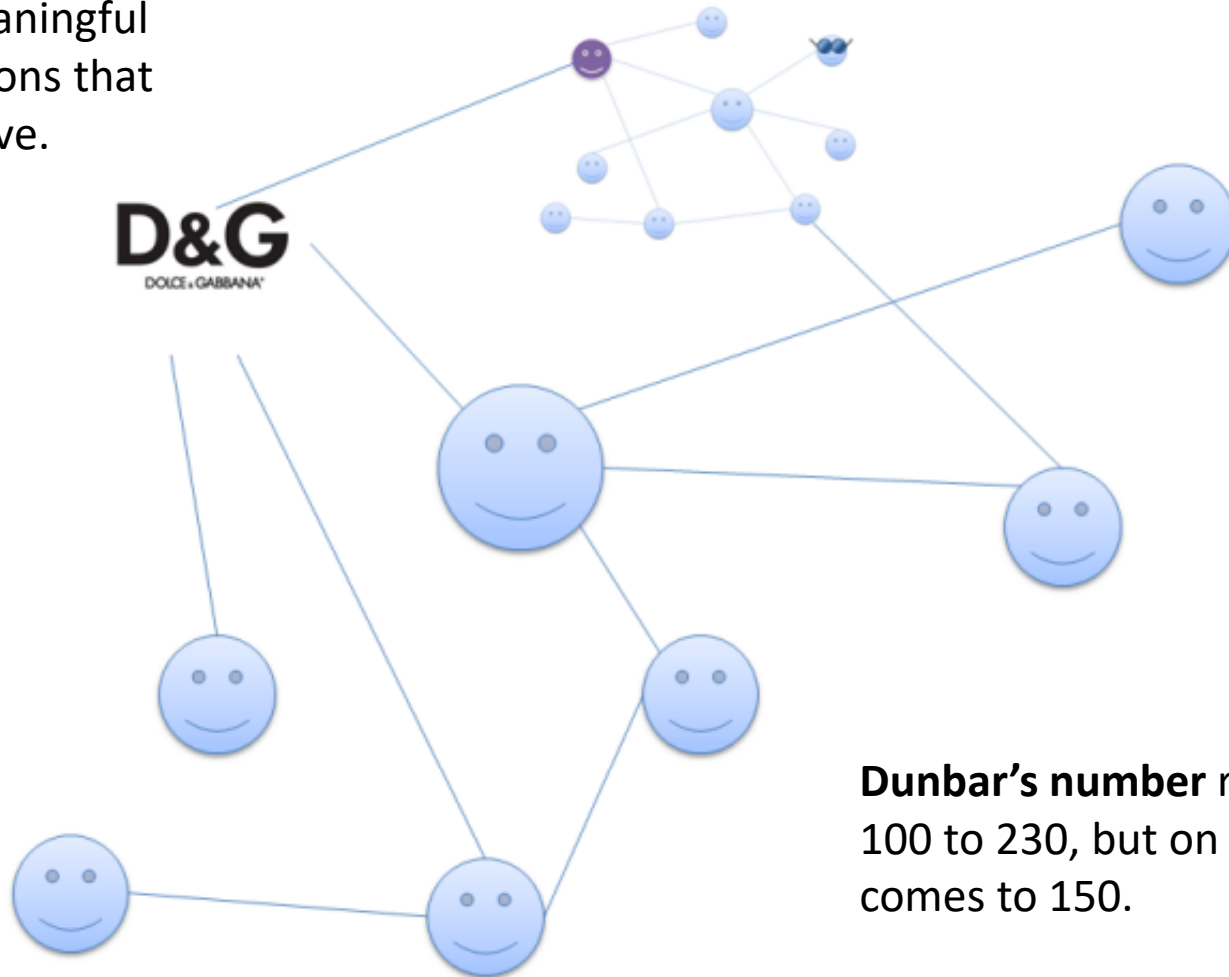


Case1: SMN



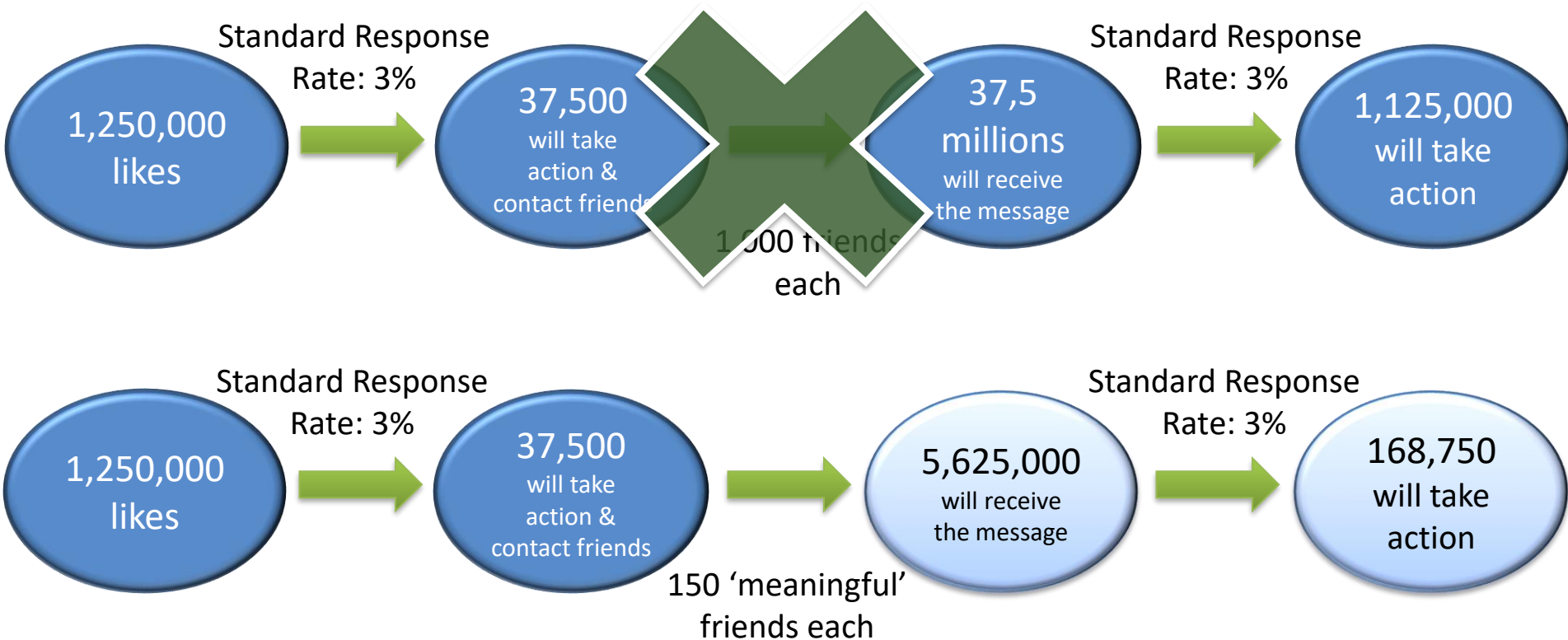
Case1: SMN

Dunbar's Number is a theoretical limit on the number of meaningful social interactions that anyone can have.



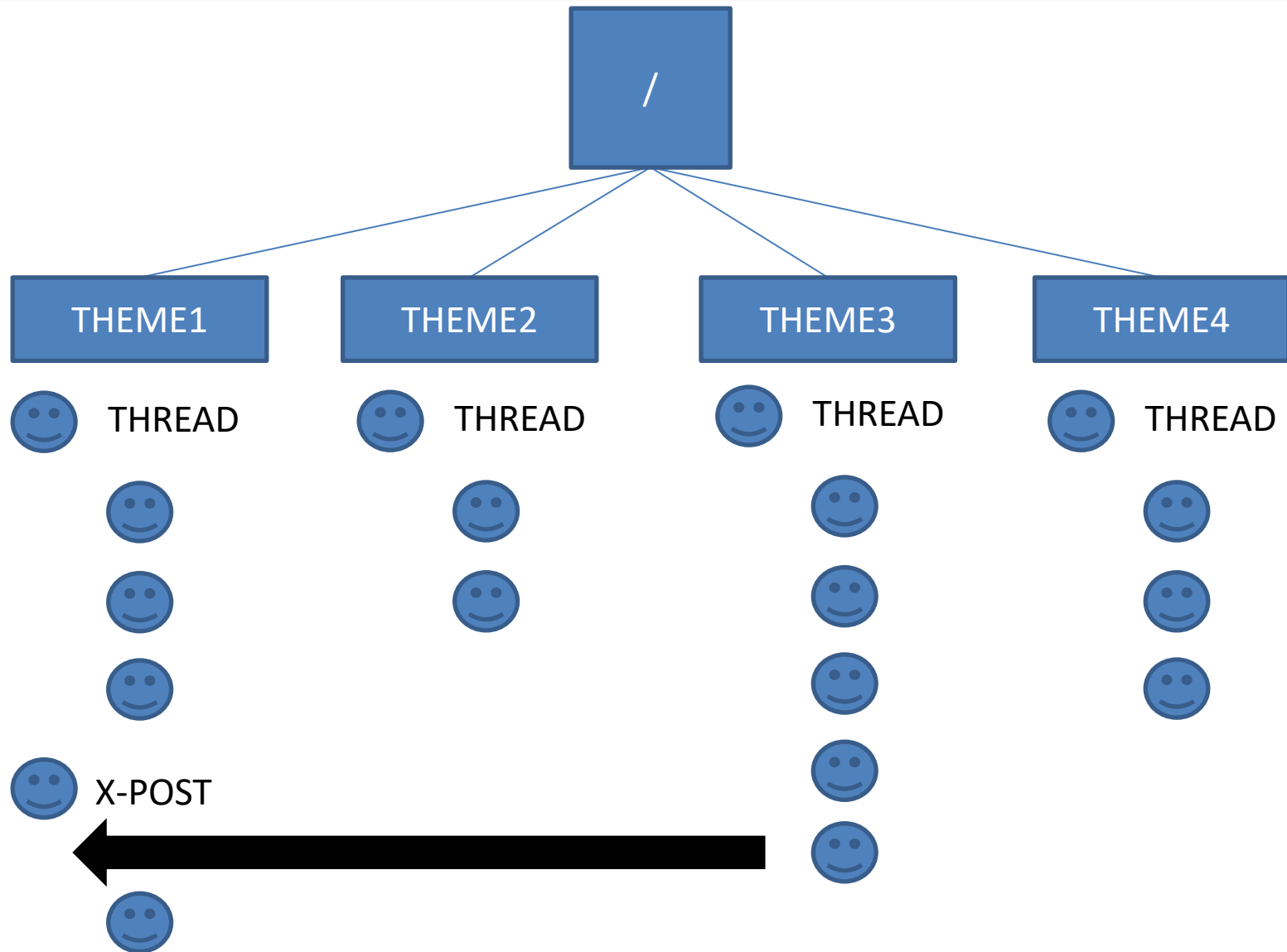
Dunbar's number ranges from 100 to 230, but on average it comes to 150.

Data Flow In Practice...



- Approx. 7.4% of the initial 'likes' might take action (if the response rate is 3%)

Case2: SM



Studying Consumer Opinion

- Millions of potential users
- Real-time
- N-way conversations
- Avoid survey/response bias

[POS]

- Reliability
- Authenticity
- Openness and transparency of platforms

[NEU]

- Potential for bias
- Poor or low quality data
- Non-standardised data formats

[NEG]

These operational aspects should be reflected on prior to conducting an analysis

Social Media Data Forms

Comments



Posts (video, photos, stories)

Actions and Events

- Likes
- Subscriptions
- Friends Requests
- Share

“On FB each month, the typical user averages 12 likes, 5 comments, and 1 share” (SocialPilot, 2023)

Data Collection Methods

1) Social Media APIs

What is an API

An application programming interface (API) is a published standard for exchanging information between different software systems or programs.

2) Web scraping (e.g. BeautifulSoup, Selenium)

3) Secondary data sources

- ❑ Kaggle
- ❑ Amazon product review database
- ❑ Data.world

Facebook Graph API

In a nutshell

The Facebook API is composed of nodes (such as a User, a Photo, a Page, a Comment), edges (such as a Page's Photos, or a Photo's Comments), and fields (such as the birthday of a User, or the name of a Page).

- Each instance of a node is an object and has a unique object ID.
- If a **Page** has a unique id of 1234, we can access the comments posted through the comments edge.
https://graph.facebook.com/1234/comments?access_token=ACCESS-TOKEN
- Access to different nodes and edges has to be approved through your Facebook Developer Account.

Facebook Graph API

https://graph.facebook.com/USER-ID?fields=id,name,email,picture&access_token=ACCESS-TOKEN

```
{
  "id": "USER-ID",
  "name": "EXAMPLE NAME",
  "email": "EXAMPLE@EMAIL.COM",
  "picture": {
    "data": {
      "height": 50,
      "is_silhouette": false,
      "url": "URL-FOR-USER-PROFILE-PICTURE",
      "width": 50
    }
  }
}
```

Data Exchange Formats

JSON

JSON stands for JavaScript Object Notation and is syntax for storing and exchanging text information. Much like XML. The equivalent JSON file is usually smaller than XML, since it contains less meta-data and formatting, this makes it faster and easier to read.

XML

```
<?xml version="1.0"?>
<modules>
  <module>
    <name>web and Social Media Analytics</name>
    <moduleLeader>Philip worrall</moduleLeader>
    <students>
      <student>
        <name>James</name>
      </student>
      <student>
        <name>Dan</name>
      </student>
    </students>
  </module>
  <module>
    <name>Business Optimisation</name>
  </module>
</modules>
```

JSON

```
{
  "modules" : [
    {
      "name": "web and Social Media Analytics",
      "moduleLeader": "Philip worrall",
      "students": [
        {
          "name": "James"
        },
        {
          "name": "Dan"
        }
      ]
    },
    {
      "name": "Business Optimisation"
    }
  ]
}
```

Twitter API

Streaming

- Messages arrive continuously in (semi) real-time.
- Must specifically stop listening or risk quota being exceeded.
- Stream can be filtered.

Historic / Timeline

- Specify the period we are interested in.
- Request a set number of tweets or all that is available.
- Query like format used to select relevant tweets.

N.B: The lowest level account enables ~2m tweets to be pulled each month.

N.B: The default endpoints allow viewing only the last 7 days of history

Example Tweet in JSON

The core object is a Tweet (tree root) containing various fields.
Each tweet is also the **parent** of other objects (User, Poll, Media, Places)

```
1 {
2   "author_id": "1579289998947676166",
3   "created_at": "2023-03-05T09:15:55.000Z",
4   "edit_history_tweet_ids": [
5     "1632550541560795136"
6   ],
7   "id": "1632550541560795136",
8   "lang": "en",
9   "possibly_sensitive": false,
10  "public_metrics": {
11    "impression_count": 0,
12    "like_count": 21,
13    "quote_count": 2,
14    "reply_count": 1,
15    "retweet_count": 12
16  },
17  "reply_settings": "everyone",
18  "text": "👹 According to the critics, the story of Henry and Sam will be
          action and emotional, the critics described this episode as one of the
          best!!\n\n P.N.: May God have mercy on us, the last time we #TheLastOfUsHBO
          #TheLastOfUs #TheLastofUs #thelastofus \nOriginal: moslhe775 https://t.co
          /JPXNWDØRGW"
19 }
```


Reddit API

- The core data model is a **submission** object
- Sub-reddit → Submission → Comments → Comment Forest

RETRIEVAL MODES

- By sub-reddit (historic)
- Monitoring (streaming)
- Search (historic)

Reddit API

Each comment and submission object store a variety of different fields

```
'author_fullname': 't2_m656hlh2',
'author_is_blocked': False,
'author_patreon_flair': False,
'author_premium': False,
'awards': [],
'banned_at_utc': None,
'banned_by': None,
'body': 'I think Elon is more worried about getting likes on twitter.',
'body_html': '<div class="md"><p>I think Elon is more worried about getting
              likes on twitter.</p>\n'
              '</div>',
'can_gild': True,
'can_mod_post': False,
'collapsed': False,
'collapsed_because_crowd_control': None,
'collapsed_reason': None,
'collapsed_reason_code': None,
'comment_type': None,
'controversiality': 0,
'created': 1676570462.0,
'created_utc': 1676570462.0,
'depth': 0,
'distinguished': None,
'downs': 0,
```

```
'edited': False,
'gilded': 0,
'gildings': {},
'id': 'j8sot60',
'is_submitter': False,
'likes': None,
'link_id': 't3_113wpzt',
'locked': False,
'mod_note': None,
'mod_reason_by': None,
'mod_reason_title': None,
'mod_reports': [],
'name': 't1_j8sot60',
'no_follow': False,
'num_reports': None,
'parent_id': 't3_113wpzt',
'permalink': '/r/news/comments/113wpzt/tesla_recalls_362758_vehicles_says_full/j8sot60/',
'removal_reason': None,
'report_reasons': None,
'saved': False,
'score': 2932,
'score_hidden': False,
'send_replies': True,
'stickied': False,
'subreddit': Subreddit(display_name='news'),
```

NB The Reddit API is one of the easier APIs to work with. Sign up for a reddit account and visit <https://www.reddit.com/prefs/apps> to create a new App and SECRET KEY

YouTube API

- **YouTube Data API** is one of several APIs
- The data model used is based on the resource type e.g.
 - Channels
 - **Videos**
 - Playlists
 - Comment Threads
 - Comments
- Query the resources we are interested in based on a search term or property
- 10,000 credits per day

YouTube API

- Find a video related to the topic of interest and identify the video ID.



- Submit requests via the API Explorer
 - https://developers.google.com/apis-explorer/?hl=en_GB#p/youtube/v3/

To access the API you can sign into the Developer Console (<https://console.cloud.google.com/>) using your University account. Create an app and subscribe to the YouTube Data API

In Summary

- Social media can encompass a variety of different websites that in some way encourage the collaboration, broadcasting of ideas/thoughts or the formation of communities or relationships.
- Some social media networks make available their data through an API (either on a free or subscription basis).
- Social media data has been used to study a variety of different phenomenon, including consumer behavior and attitudes.
- Analysis of social media data requires specialist approaches due to its unstructured nature and volume.
- Various advantages and disadvantages exist with respect to using social media data for the purposes of collecting consumer insight.

End