



# Social Media Personality Profiling

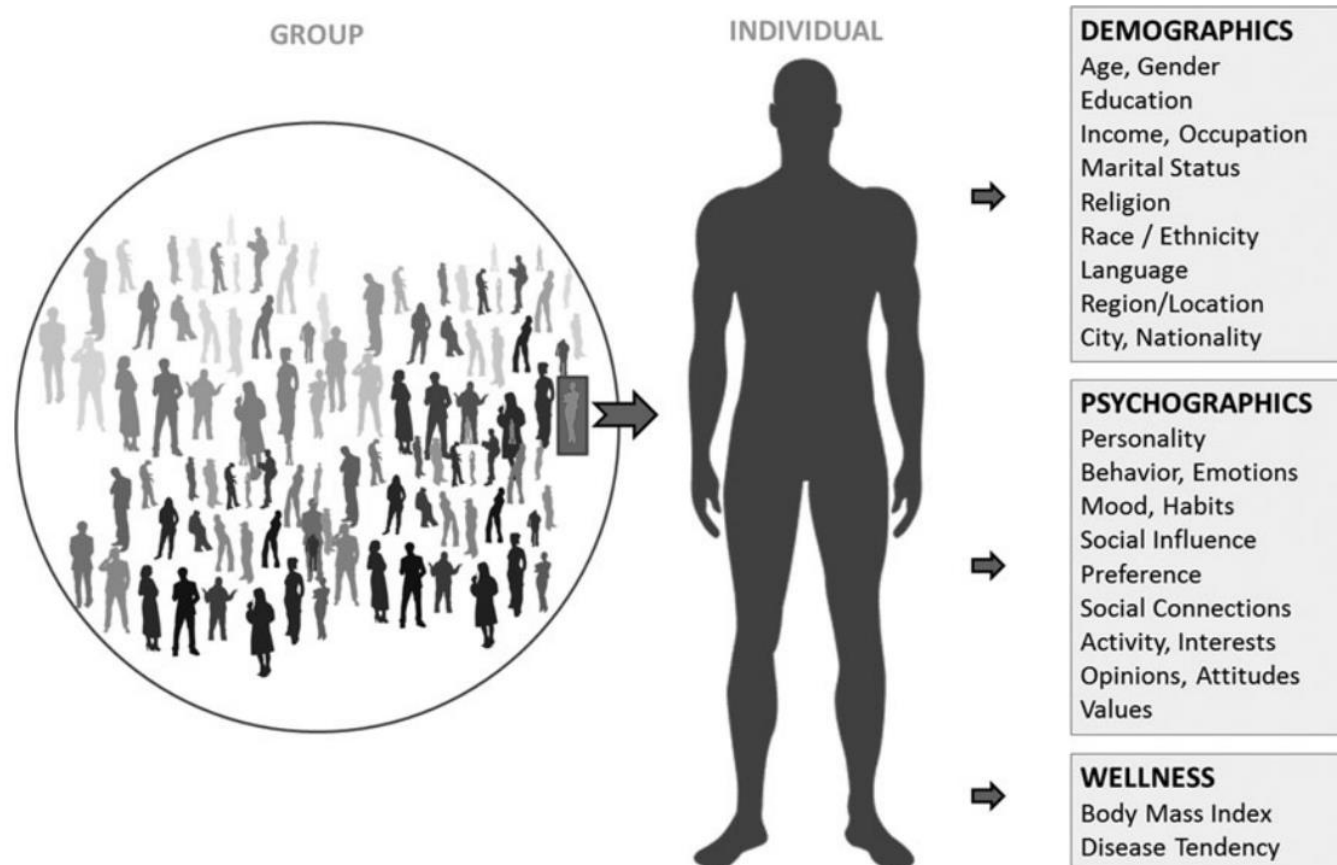
Abbas Maazallahi

[a.maazallahi@ut.ac.ir](mailto:a.maazallahi@ut.ac.ir)

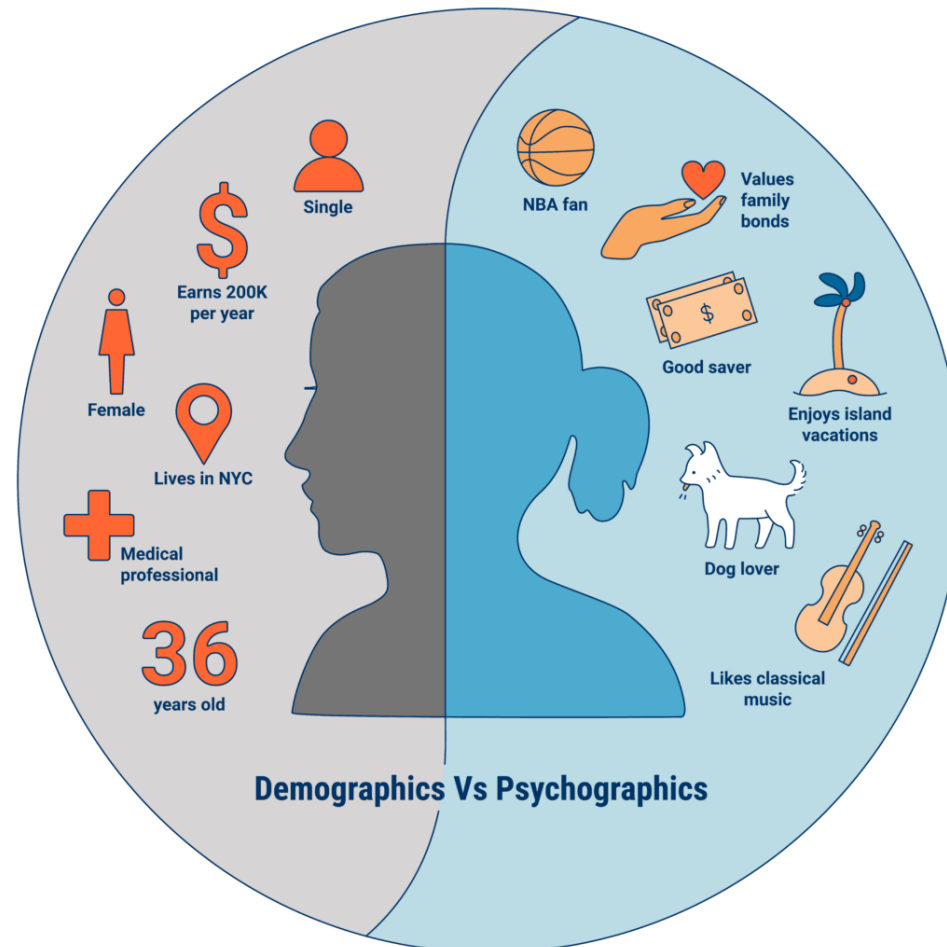
[https://t.me/a\\_maazallahi](https://t.me/a_maazallahi)

8/20/2020

# Taxonomy of social profiling

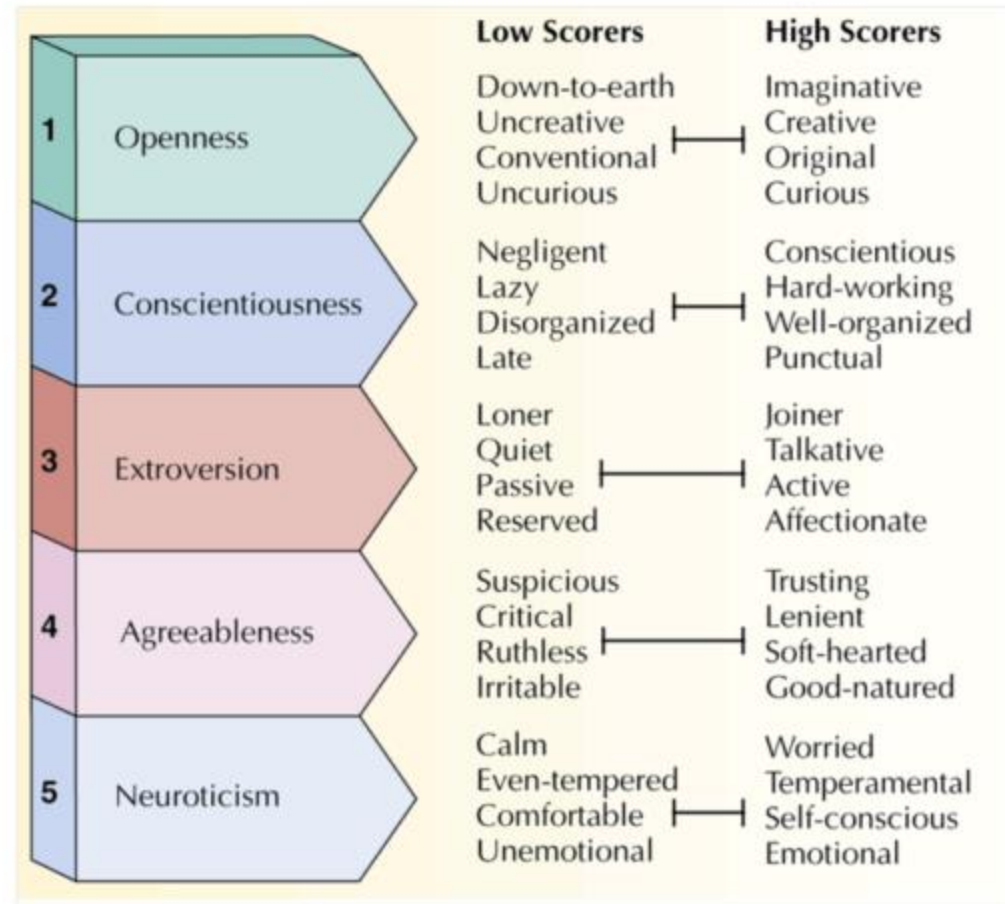


# Demographics vs Psychographics



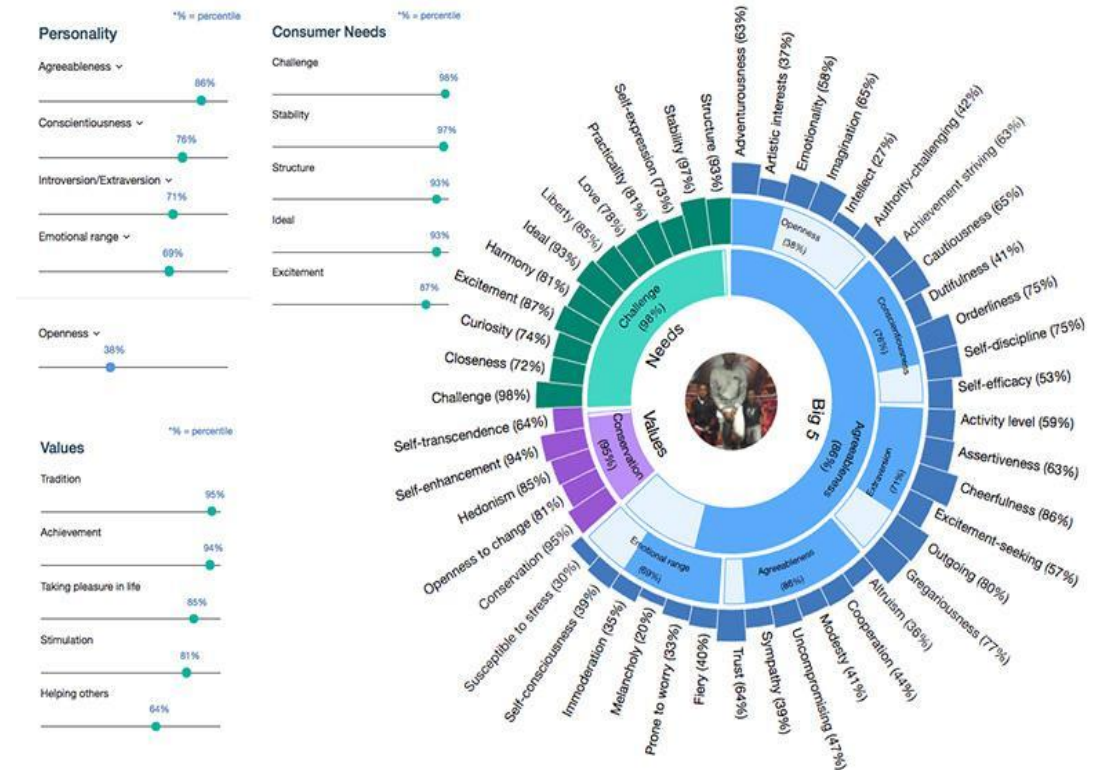
# OCEAN model of personality (psychographic profile)

<https://www.cbinsights.com/research/what-is-psychographics/>



# IBM's Personality Insights tool

<https://www.welcome.ai/tech/personalization/ibm-watson-personality-insights>



# Cognitive Ability

<https://practicereasoningtests.com/cognitive-ability-tests/>

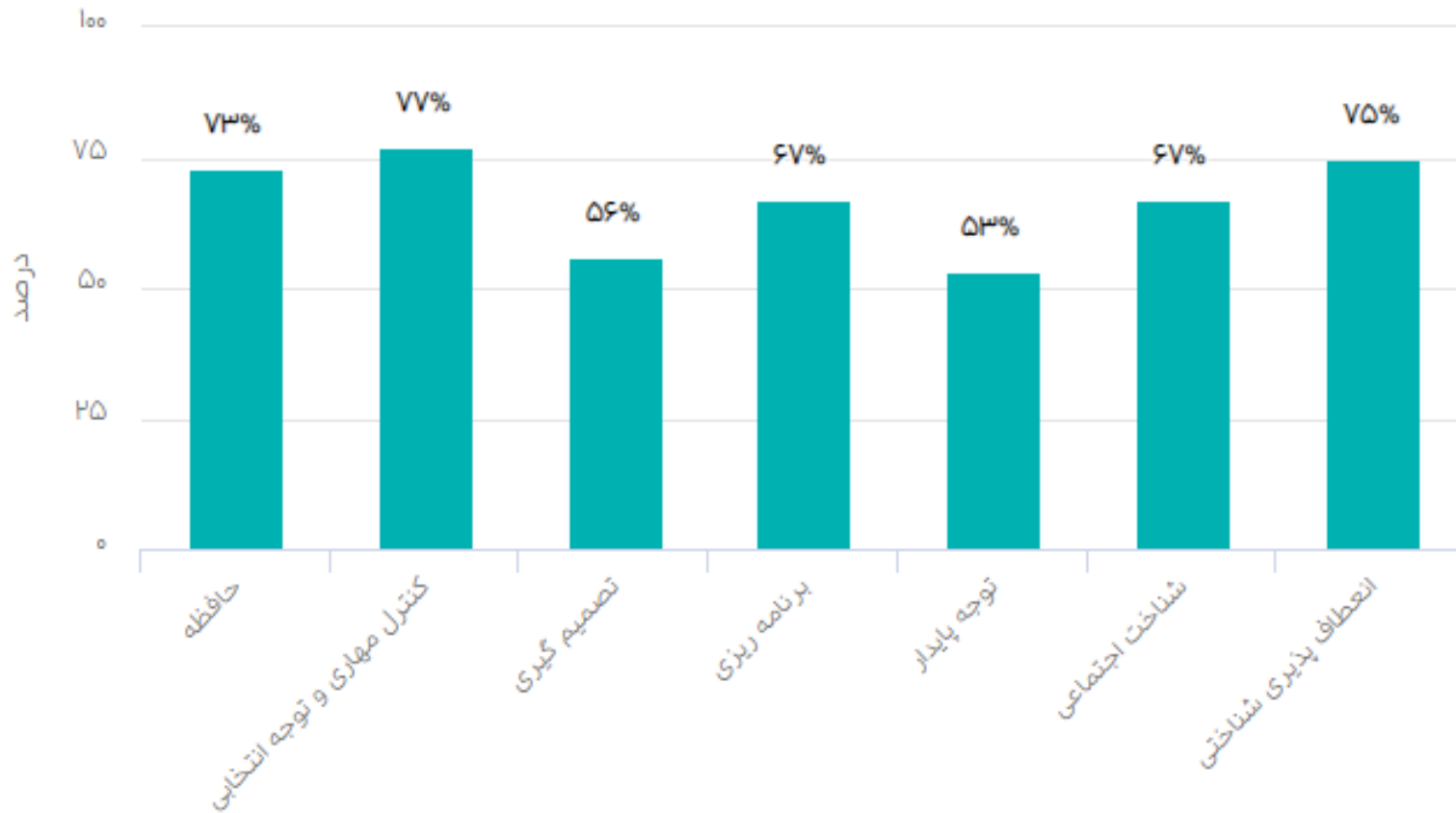
The most popular cognitive ability assessment that is built on the General Ability theory and has been in use for years now is the Intelligence Quotient (IQ) test

- Verbal (linguistic): lexical skills, formal speech, verbal argument, and creative writing.
- Body (kinesthetic/movement): body language, physical gestures, creative dance, physical exercise, drama.
- Musical (rhythmic): music performance, singing, musical composition, and rhythmic patterns.
- Logic (mathematic): numerical aptitude, problem solving, and deciphering codes, abstract symbols, and formulae.
- Visual (spatial): Developing and recognising patterns and designs, painting, drawing, active, imagination, sculpture, and colour schemes.
- Interpersonal (relationships with others): The person-to-person communication, empathy practices, group projects, collaboration skills, receiving and giving feedback.
- Intrapersonal (self-understanding and insight): thinking strategies, emotional processing, knowing yourself, higher order reasoning, focusing and concentration.

Also, *Creativity* and *Memory* are sometimes added as separate intelligences, making up the eighth and the ninth intelligence categories under the Gardner model.

# Cognitive Ability

نمودار توانمندی های شناختی شما



انعطاف پذیری شناختی

شناخت اجتماعی

توجه پایدار

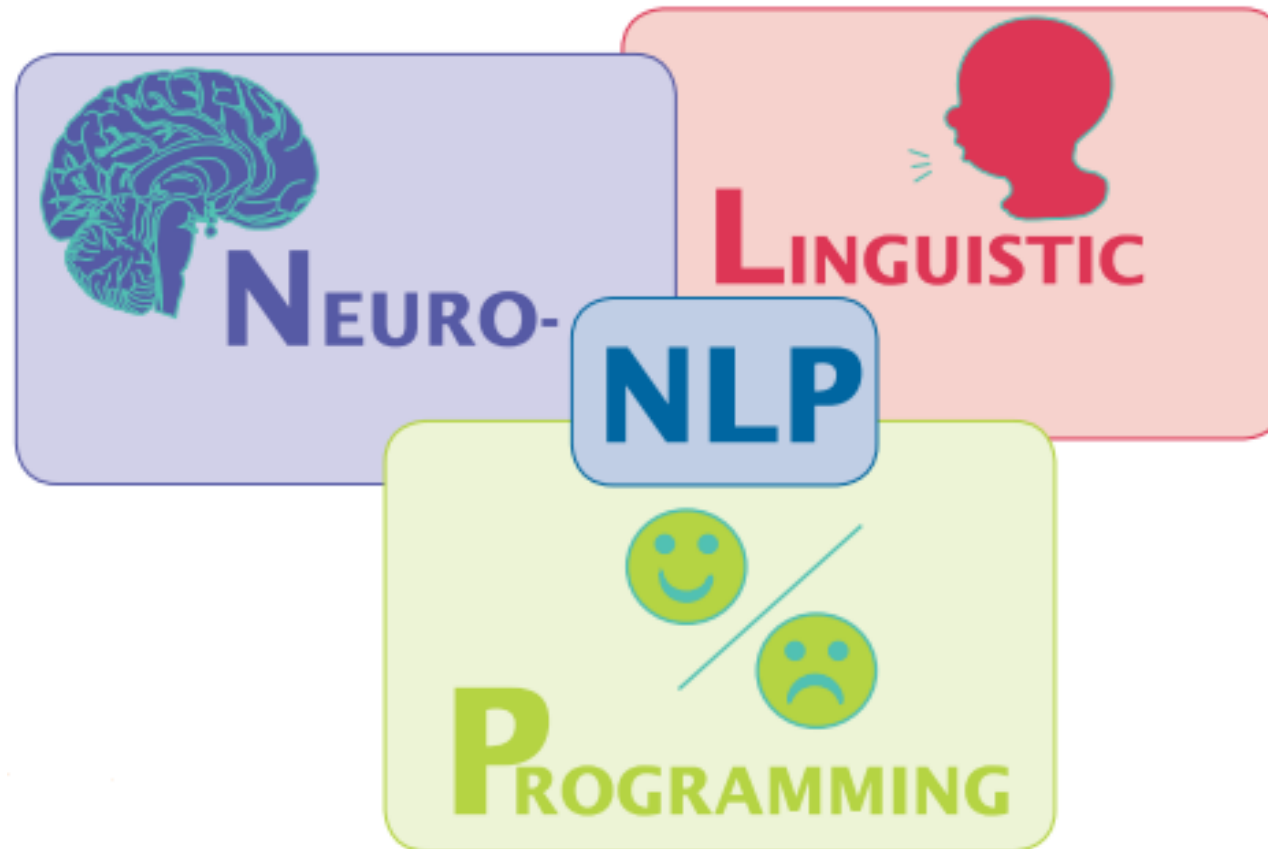
برنامه ریزی

تصمیم گیری

کنترل مهارت و توجه انتخابی

حافظه

## دسته بندی پروژه‌های مرتبط با ساخت پروفایل کاربران



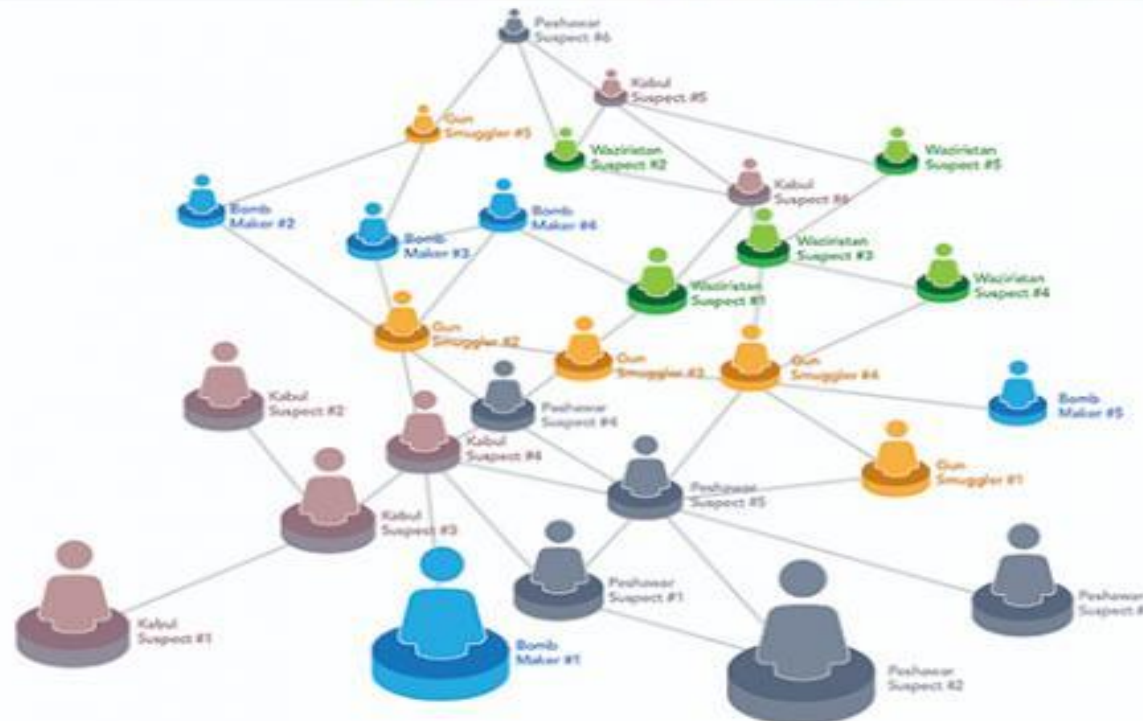


## دسته بندی پروژه‌های مرتبط با ساخت پروفایل کاربران



# دسته بندی پروژه‌های مرتبط با ساخت پروفایل کاربران

## SOCIAL NETWORK ANALYSIS



دسته بندی پروژه‌های مرتبط با ساخت پروفایل کاربران



# پردازش متن در شبکه‌های اجتماعی

متن‌های مربوط به کاربران حاوی اطلاعات ارزشمندی می‌باشد که امکان تحلیل و پردازش دارند.

به طور کلی می‌توان متن‌های مختلفی را منتسب به کاربر کرد:

- ▶ نام کاربری
- ▶ نام مستعار
- ▶ بیوگرافی
- ▶ متن‌های منتشر شده در قالب توییت، پست و پیام
- ▶ کامنت‌های گذاشته شده بر روی پست‌های خود و دیگر کاربران
- ▶ کامنت‌های دریافتی از دیگر کاربران
- ▶ هشتک‌های استفاده شده
- ▶ کاربران منشن شده
- ▶ لینک‌های URL استفاده شده
- ▶ emojiهای قرار داده شد
- ▶ ...

# پردازش متن در شبکه‌های اجتماعی

اطلاعات مختلفی را از طریق پردازش متن در شبکه‌های اجتماعی می‌توان استخراج کرد:

- ▶ زبان کاربر (Classifier)
- ▶ گویش کاربر (Classifier)
- ▶ علائق کاربر (Classifier)
- ▶ شناسایی سن و جنسیت کاربر (Classifier or Lookup table)
- ▶ شناسایی احساس و جهت گیری کاربر پیرامون موضوعات مختلف (Classifier)
- ▶ یافتن نگرش کاربر حول موضوعات مختلف (Classifier)
- ▶ اطلاعات مربوط به شغل و حرفه و توانی‌های کاربر (Classifier)
- ▶ ارتباطات و کیفیت ارتباط با دیگر کاربران (Link Prediction)
- ▶ ...

# facebookresearch/fastText

fastText is a library for efficient learning of word representations and sentence classification.

Models:

- ◆ Recent state-of-the-art English word vectors.
- ◆ Word vectors for 157 languages trained on Wikipedia and Crawl.
- ◆ Models for language identification and various supervised tasks.

The logo for fastText, with 'fast' in a red, italicized sans-serif font and 'Text' in a blue, bold sans-serif font.

# spaCy: Industrial-strength NLP

- ▶ spaCy is a library for advanced Natural Language Processing
- ▶ built on the very latest research,
- ▶ spaCy comes with pretrained statistical models and word vectors
- ▶ supports tokenization for 50+ languages
- ▶ convolutional neural network models for tagging, parsing and named entity recognition



# Tweet2Vec



Challenges of social media text:

- ▶ Informal language, spelling errors, abbreviations, and special characters are all commonplace in these posts,
- ▶ leading to a prohibitively large vocabulary size for word-level approaches.

Tweet2vec is a character composition model, which finds vector-space representations of whole tweets by learning complex, non-local dependencies in character sequences. The proposed model outperforms a word-level baseline at predicting user-annotated hashtags associated with the posts, doing significantly better when the input contains many out-of-vocabulary words or unusual character sequences. Our tweet2vec encoder is publicly available.



# پیش بینی جنسیت و سن از طریق نام کاربری افراد

```
>>> import gender_guesser.detector as gender
>>> d = gender.Detector()
>>> print(d.get_gender(u"Bob"))
male
>>> print(d.get_gender(u"Sally"))
female
>>> print(d.get_gender(u"Pauley")) # should be androgynous
andy
```

amos	rare	۱	۱	عاموس	3888
aye	rare	۱		عایشه	3889
abad	normal		۱	عباد	3890
abdullah	rare		۱	عبداللہ	3891
abbas	common		۱	عباس	3892
abdulamin	normal		۱	عبدالامین	3893
abdulbas	rare		۱	عبدالباسط	3894

- نام هر فرد می تواند با دقت بالایی جنسیت را تشخیص دهد و علاوه بر آن با استفاده از توزیع نام گذاری در سال های مختلف می توان به برآوردی نه چندان دقیق از سن هر فرد دست یافت. به طور مثال نام های ستایش، امیرعلی، مهدیار و ... بیشتر در سال های اخیر مورد استفاده قرار گرفته است. علاوه بر این میتوان با استفاده از نام، تا حدود خوبی قومیت و زادگاه و حتی ریشه های مذهبی و دینی هر خانواده را استخراج کرد.
- نکته ی مهم در این موضوع استخراج نام فرد از شبکه های اجتماعی می باشد. با تحلیل نام کاربری، و بیوگرافی هر فرد و استفاده از تکنیک های NER می توان نام مربوط به فرد را تشخیص داد.

# Predict Race and Ethnicity From Name

<https://github.com/appeler/ethnicolr>

Predict Race and Ethnicity Based on the Sequence of Characters in a Name



# Language Prediction

<https://github.com/nickdavidhaynes/spacy-cld>

<https://github.com/indix/whatthelang>

Enriching Word Vectors with Subword Information

[1] P. Bojanowski\*, E. Grave\*, A. Joulin, T. Mikolov, Enriching Word Vectors with Subword Information

Bag of Tricks for Efficient Text Classification

[2] A. Joulin, E. Grave, P. Bojanowski, T. Mikolov, Bag of Tricks for Efficient Text Classification

FastText.zip: Compressing text classification models

[3] A. Joulin, E. Grave, P. Bojanowski, M. Douze, H. Jégou, T. Mikolov, FastText.zip: Compressing text classification models

# Twitter Sentiment Analysis

<https://github.com/abdufatir/twitter-sentiment-analysis>

Sentiment analysis on tweets using:

**Naive Bayes**

**Maximum Entropy**

**Decision Tree**

**Random Forest**

**XGBoost**

**SVM**

**Multi-Layer Perceptron**

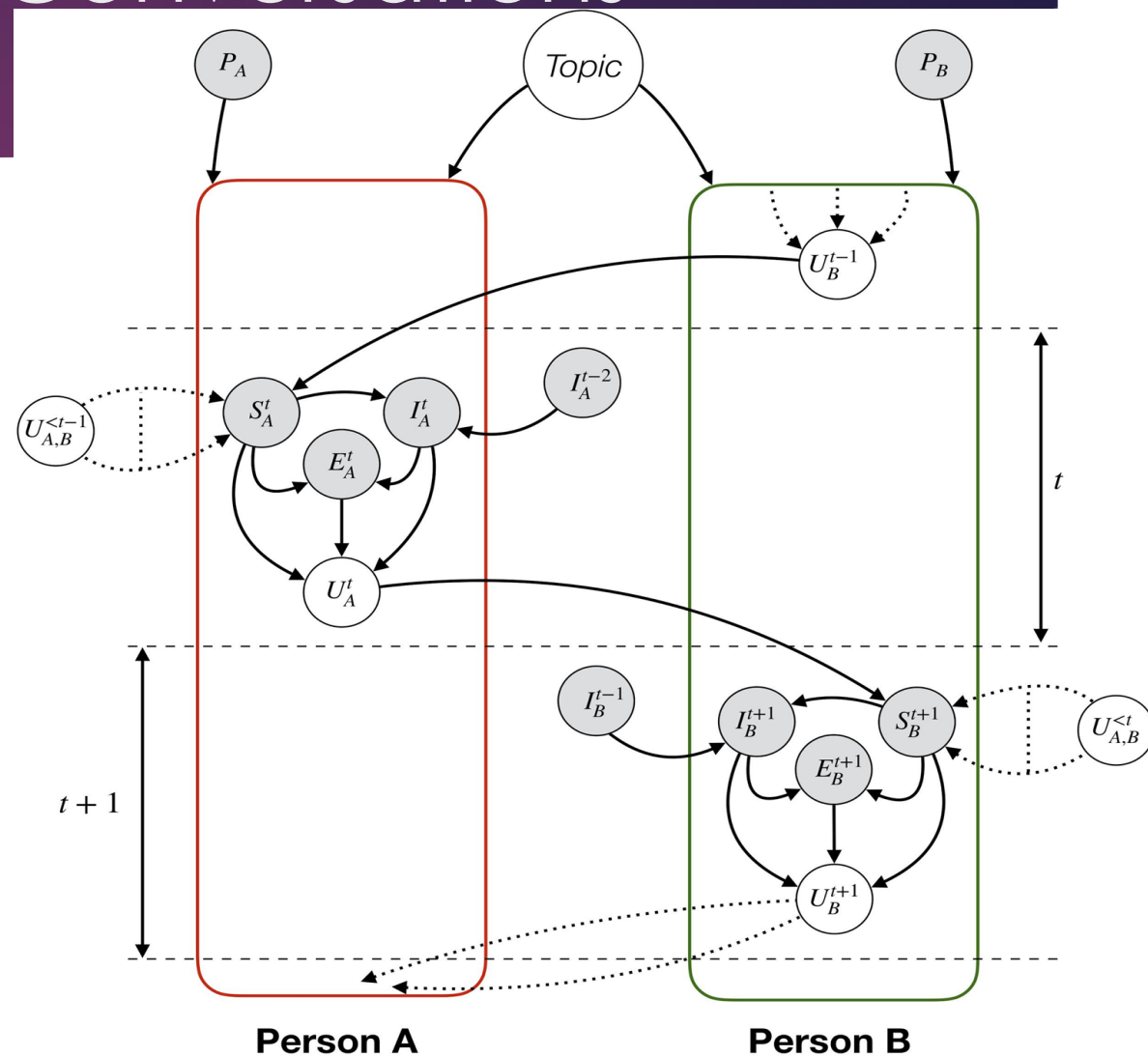
**Reccurent Neural Networks**

**Convolutional Neural Networks**

**Majority Vote Ensemble**

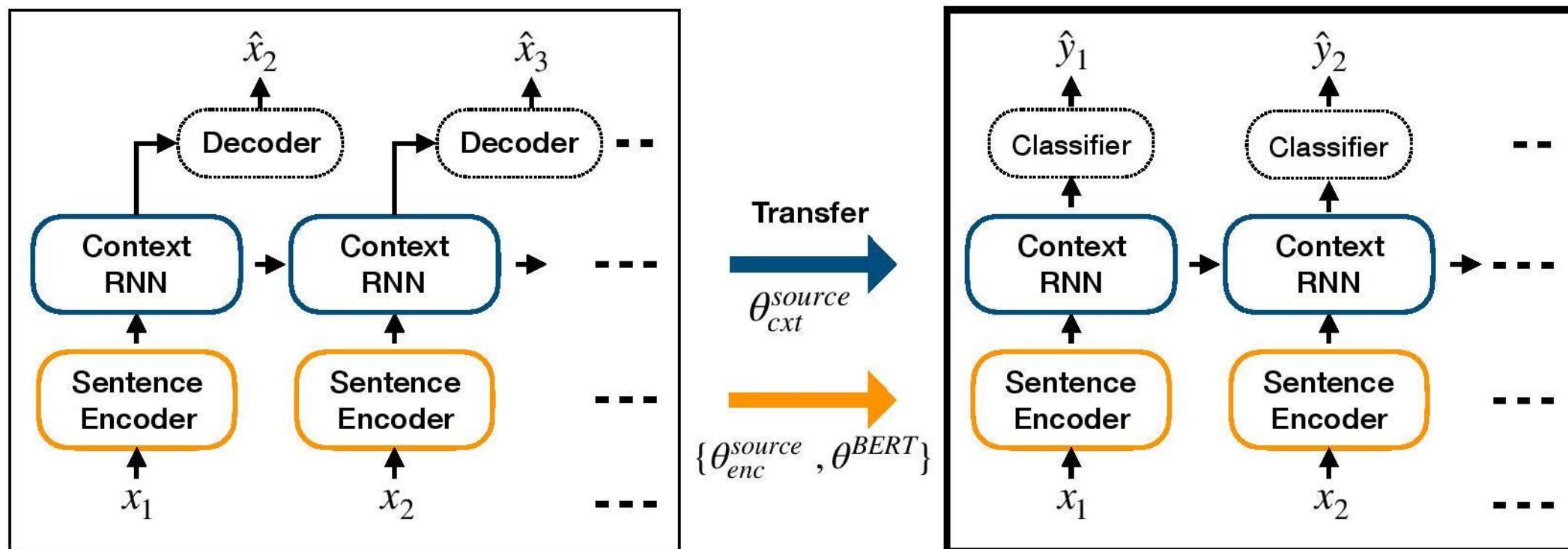
# Emotion Recognition in Conversations

<https://github.com/SenticNet/conv-emotion>



# Emotion Recognition in Conversations

<https://arxiv.org/pdf/1910.04980.pdf>



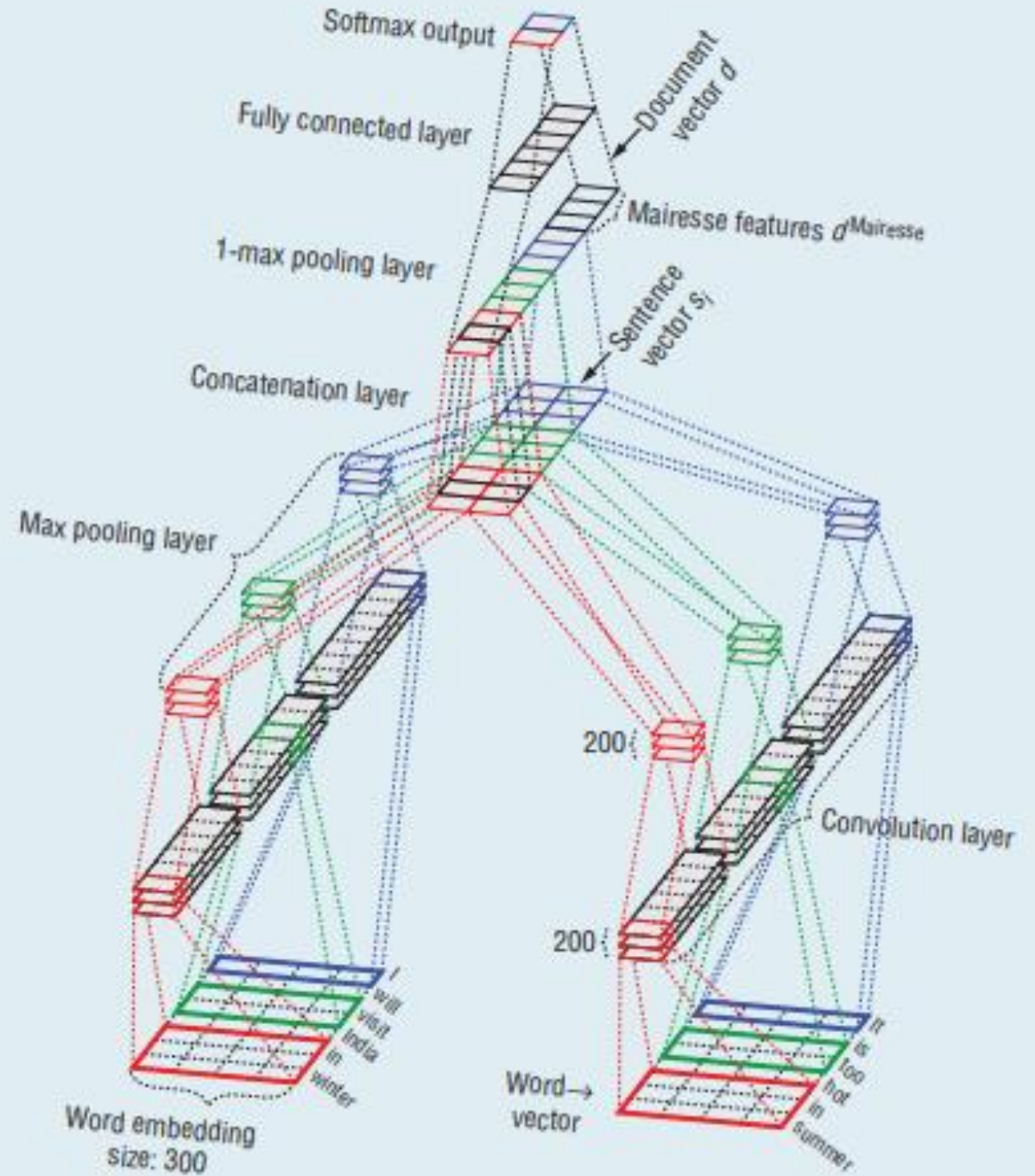
*source task:*  
**Conversation Modeling**

*target task:* **Emotion Recognition in Conversations**



# Deep Learning- Based Document Modeling for Personality Detection from Text

<https://github.com/SenticNet/personality-detection>



# پردازش تصویر در شبکه‌های اجتماعی





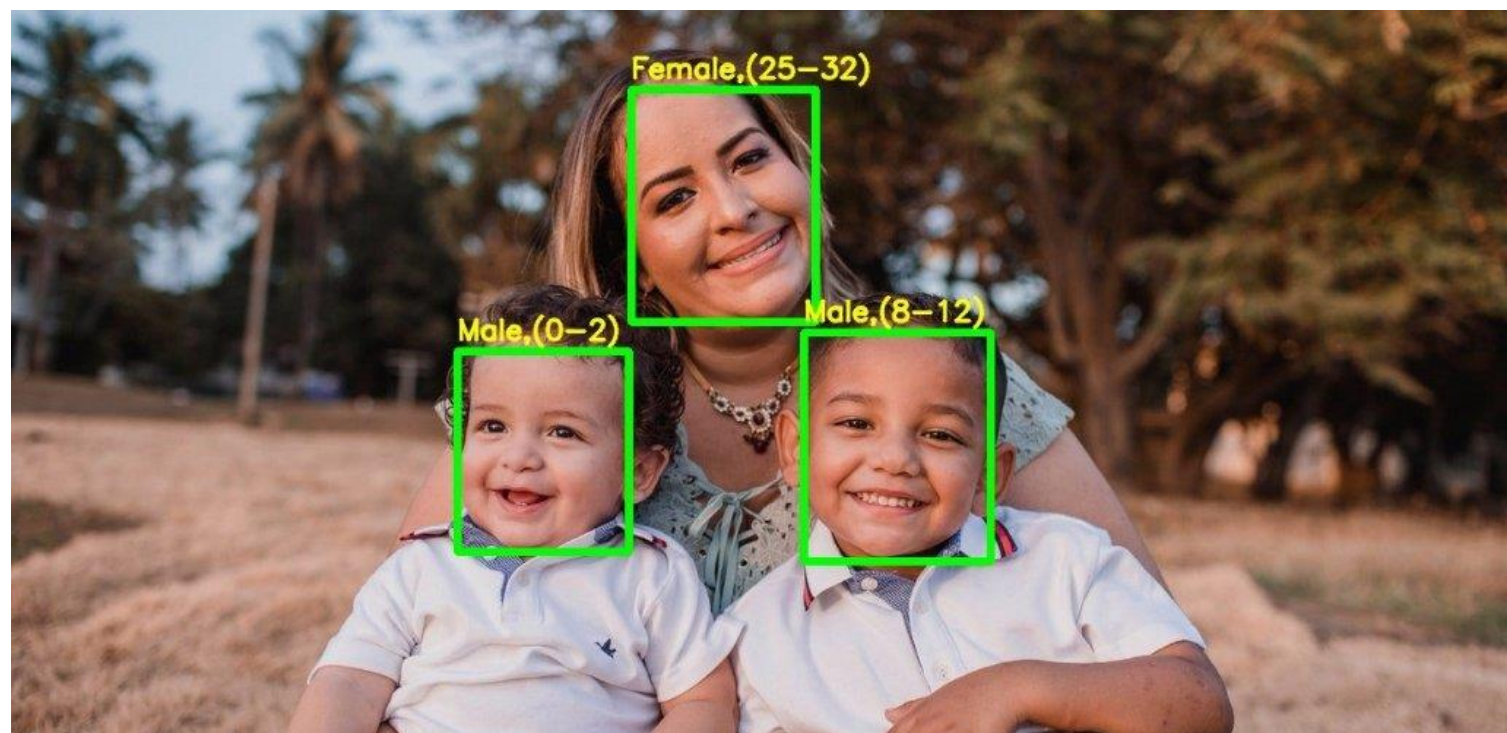
# پروژه‌های پیشنهادی در پردازش تصویر کاربران

- ▶ استخراج سن و جنسیت افراد (Age Gender Prediction)
- ▶ استخراج حالات روحی افراد (Facial Expression Recognition)
- ▶ استخراج ویژگی‌های ظاهری چهره‌ی افراد
- ▶ شناسایی نژاد افراد (Race and Ethnicity Prediction)
- ▶ استخراج شاخص تراکم بدن (Body Mass Index)
- ▶ شناسایی غذاهای مورد علاقه کاربران و کالری مصرفی

# Age Gender Estimation

<https://github.com/GilLevi/AgeGenderDeepLearning>

<https://github.com/yu4u/age-gender-estimation>



# Facial Expression Recognition

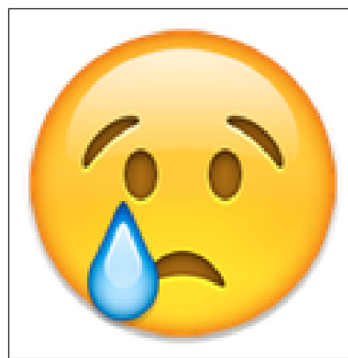
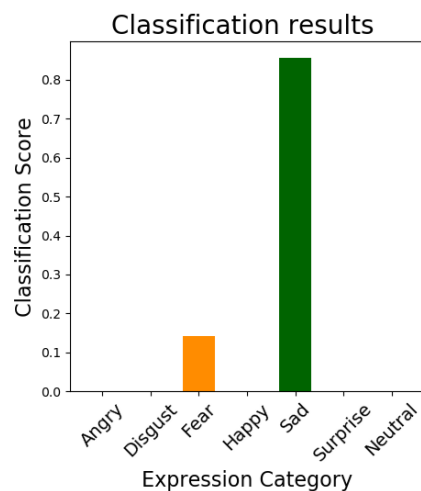
<https://github.com/WuJie1010/Facial-Expression-Recognition.Pytorch>

<https://github.com/thoughtworksarts/EmoPy>

<https://github.com/serengil/tensorflow-101>



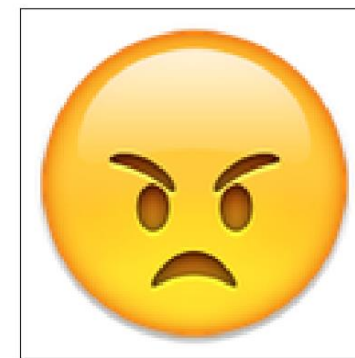
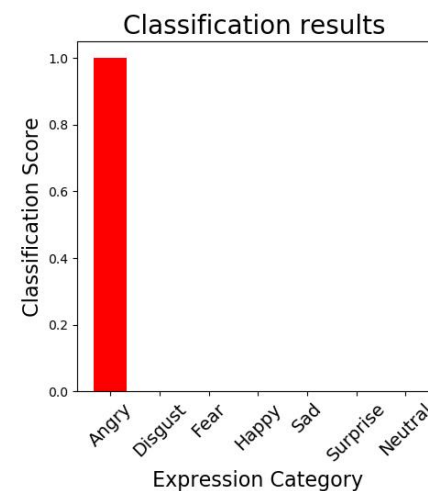
Input Image



Emoji Expression



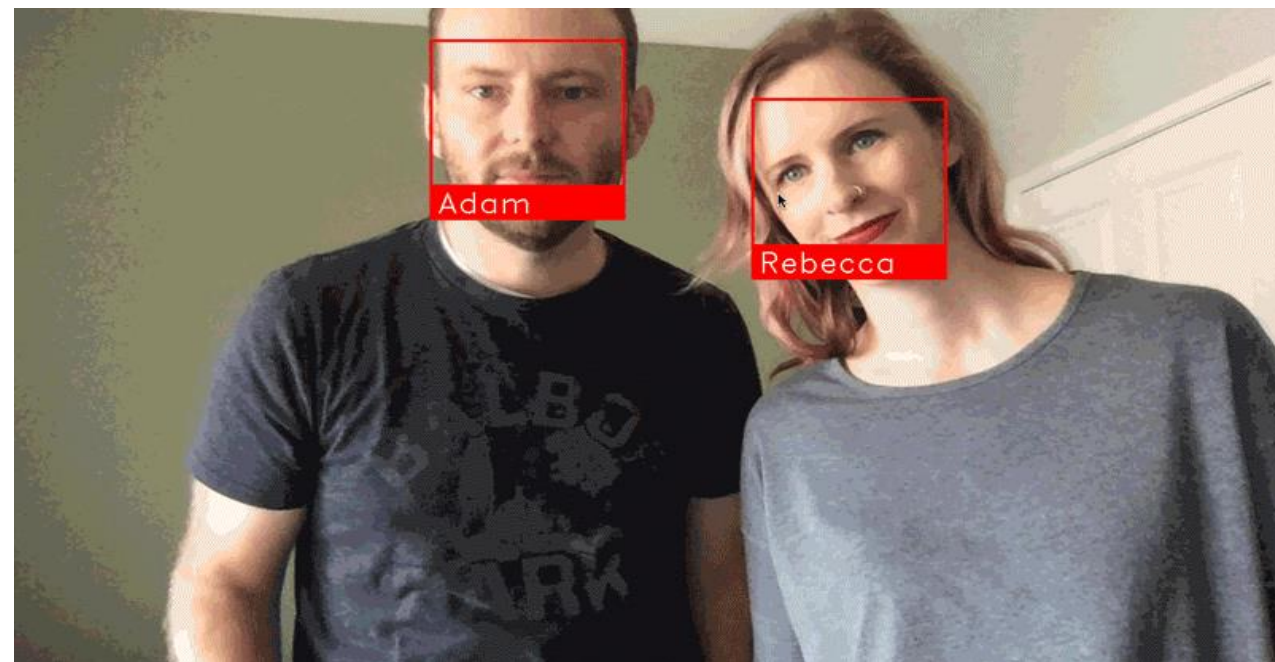
Input Image



Emoji Expression

# Face Recognition

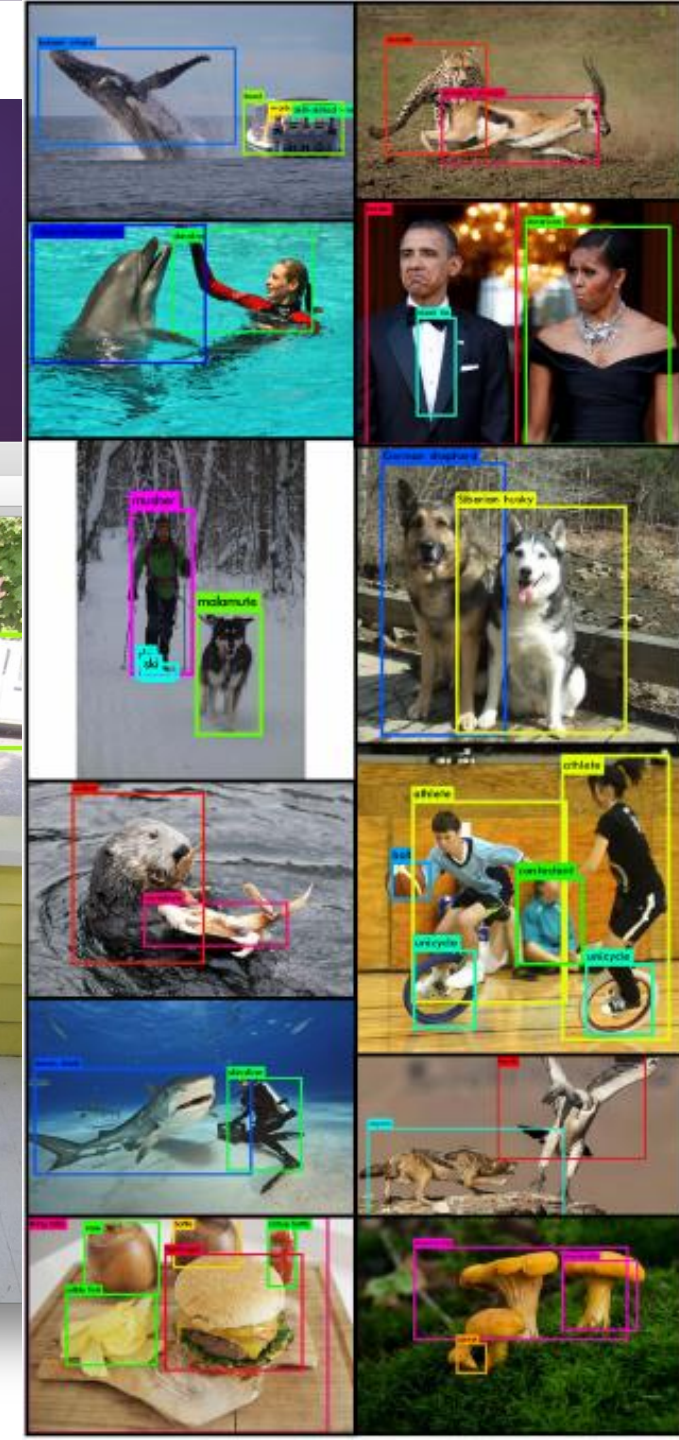
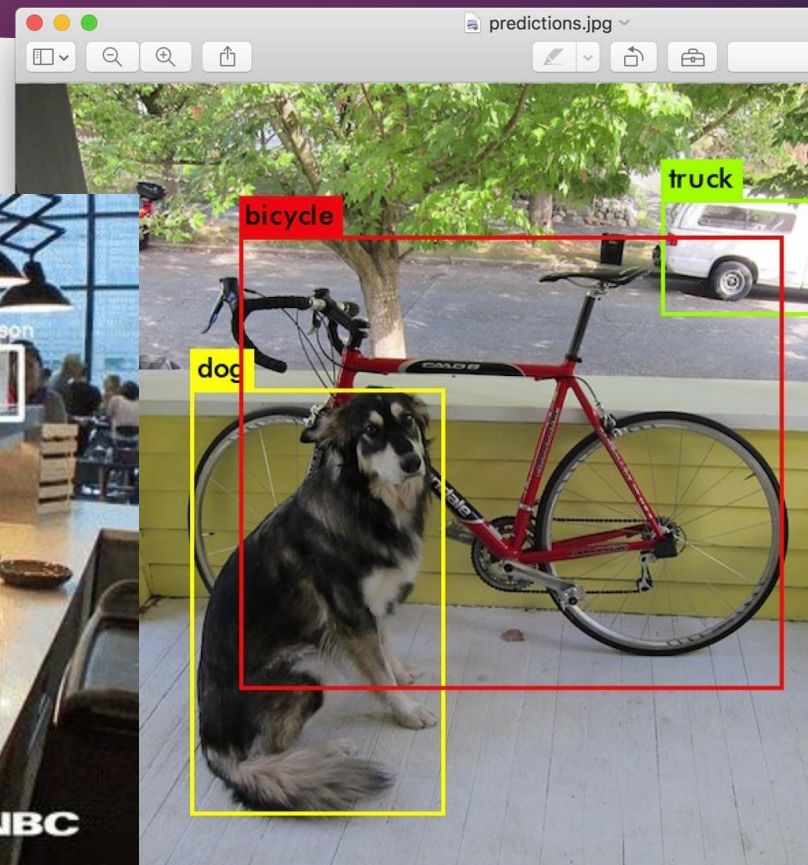
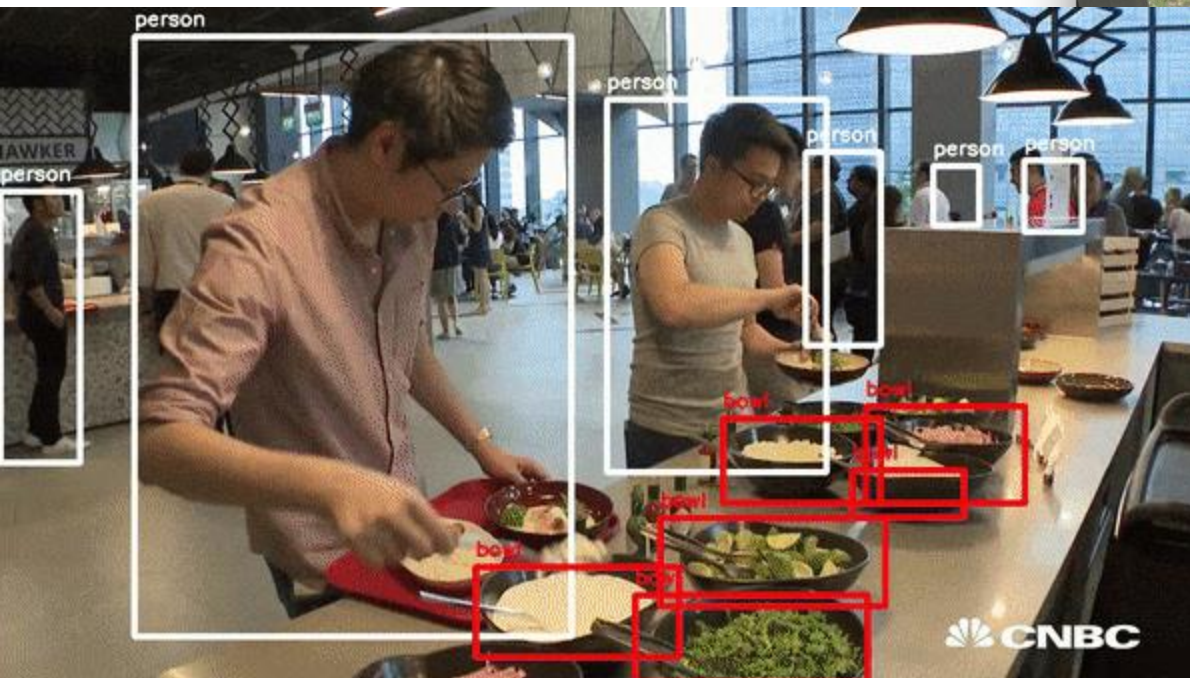
[https://github.com/ageitgey/face\\_recognition](https://github.com/ageitgey/face_recognition)





# Real-time object detection and classification

<https://github.com/thtrieu/darkflow>

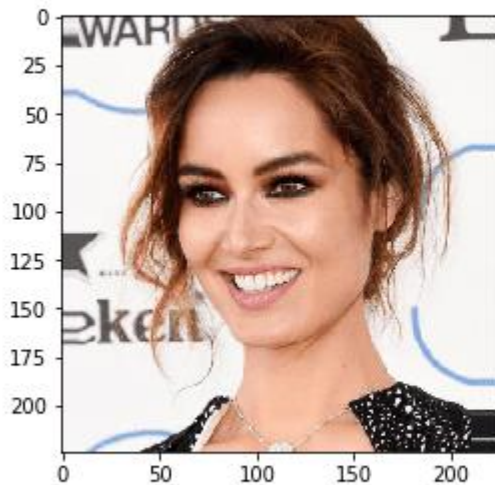




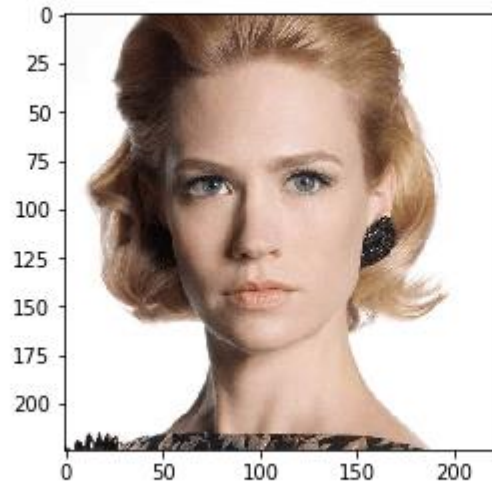
# Beauty Score Prediction

<https://github.com/ustcqidi/BeautyPredict>

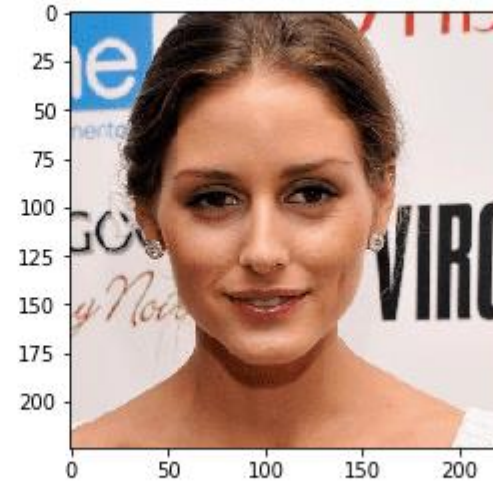
Bérénice Marlohe ( Woman )  
Beauty score: 111.71



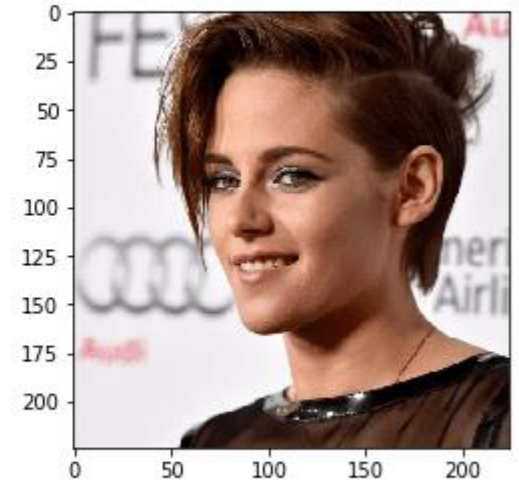
January Jones ( Woman )  
Beauty score: 105.07



Olivia Palermo ( Woman )  
Beauty score: 101.13



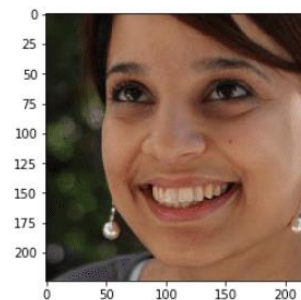
Kristen Stewart ( Woman )  
Beauty score: 100.29



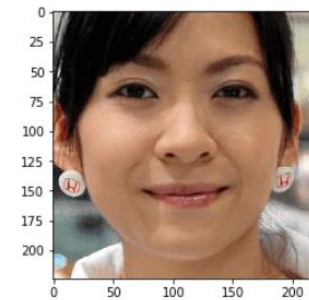
# Race and Ethnicity Prediction From Image

<https://github.com/HectorAnadon/Face-expression-and-ethnic-recognition>

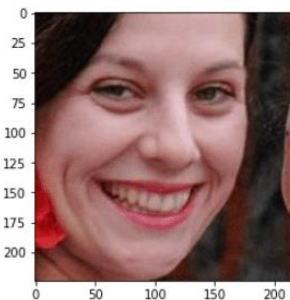
Actual: Latino\_Hispanic  
Predicted: Latino\_Hispanic



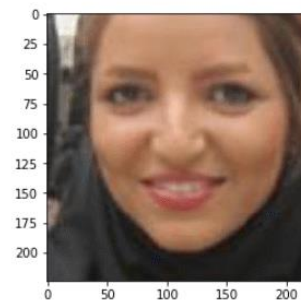
Actual: Asian  
Predicted: Asian



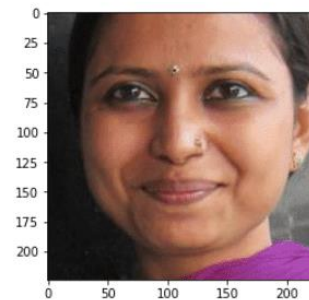
Actual: White  
Predicted: White



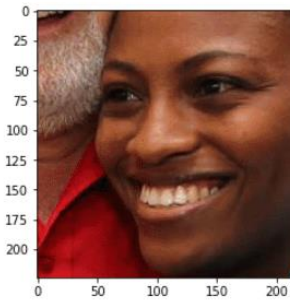
Actual: Middle Eastern  
Predicted: Middle Eastern



Actual: Indian  
Predicted: Indian



Actual: Black  
Predicted: Black



# Attractiveness Score Prediction

<https://github.com/LiuXiaolong19920720/predict-facial-attractiveness>



5.4782608



5.0869565



4.8901098

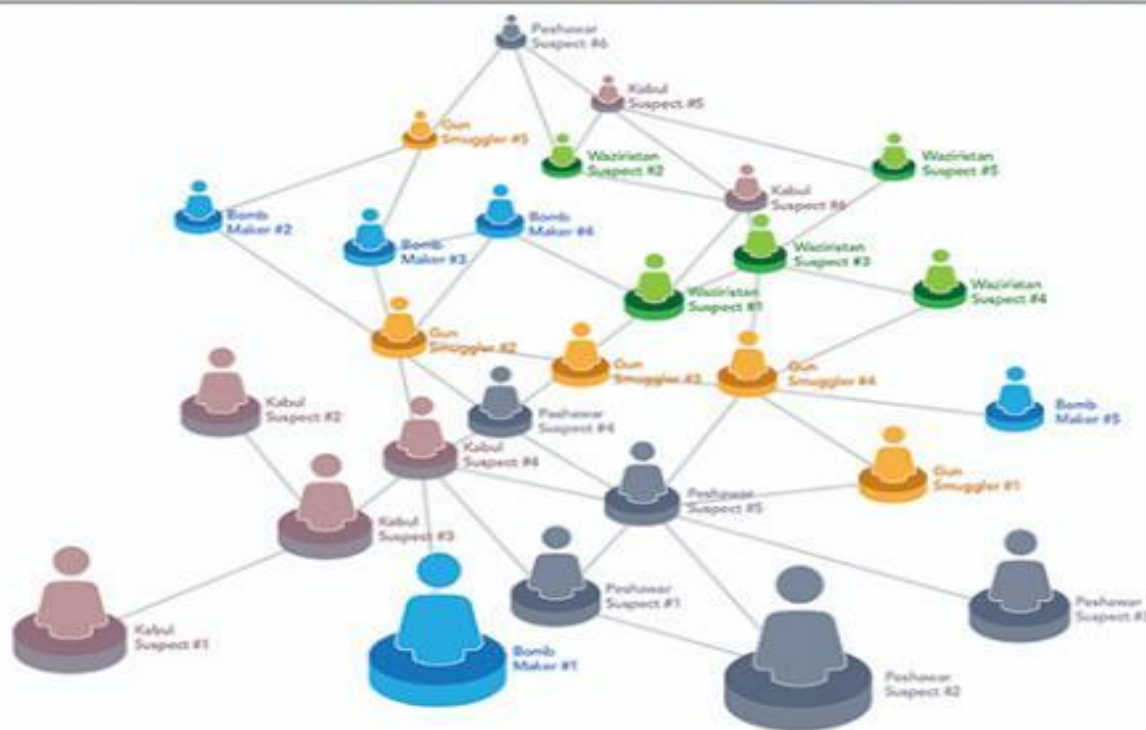


4.7553191



# تحلیل ارتباطات در شبکه‌های اجتماعی

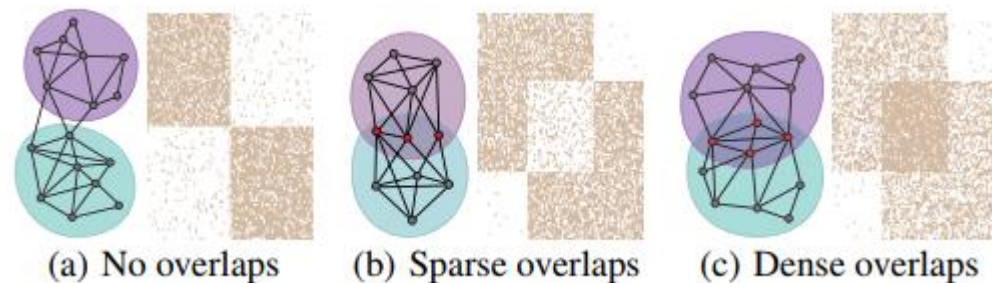
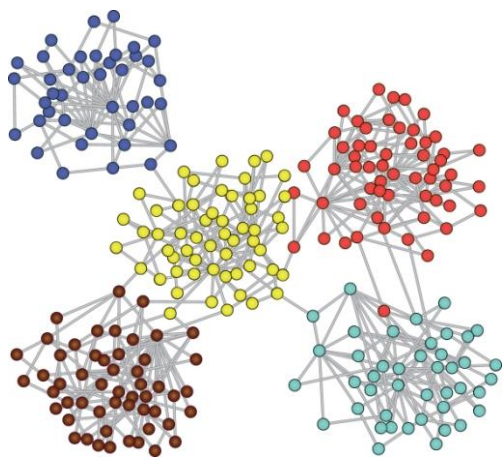
## SOCIAL NETWORK ANALYSIS



# تحلیل ارتباطات در شبکه‌های اجتماعی

شناسایی انجمن در شبکه‌های اجتماعی

الگوریتم‌های مختلفی در این مورد وجود دارد که هر کدام بسته به ویژگی‌های گراف کاربردهای مختلفی دارند. بعضی از آنها برای شناسایی انجمن‌های دارای همپوشانی قوی کاربرد دارند. بعضی از آنها برای شناسایی انجمن‌ها در گراف‌های تنک کاربرد دارند و ...



# NetworkX

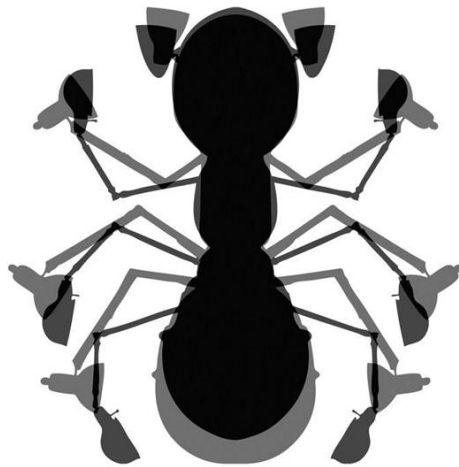
NetworkX is a Python package for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks.

## Features

- Data structures for graphs, digraphs, and multigraphs
- Many standard graph algorithms
- Network structure and analysis measures
- Generators for classic graphs, random graphs, and synthetic networks
- Nodes can be "anything" (e.g., text, images, XML records)
- Edges can hold arbitrary data (e.g., weights, time-series)
- Open source [3-clause BSD license](#)
- Well tested with over 90% code coverage
- Additional benefits from Python include fast prototyping, easy to teach, and multi-platform

# KarateClub

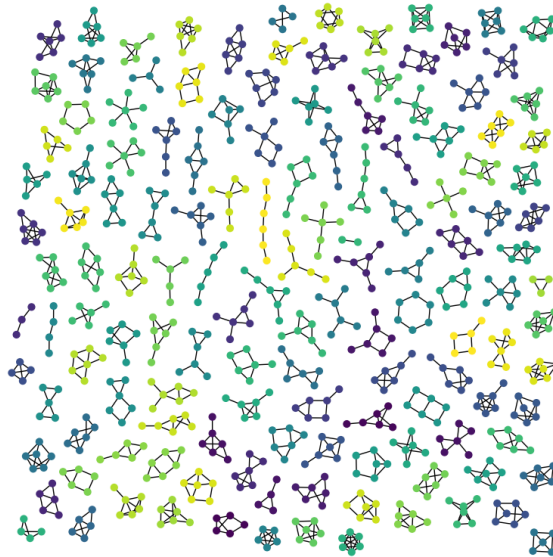
A general purpose community detection and network embedding library for research built on NetworkX.



# KARATE CLUB

# Graph Classification

یکی دیگر از رویکردهای تحلیل گراف در شبکه‌های اجتماعی اجرای روش‌های دسته‌بندی وابسته به ارتباطات در شبکه می‌باشد. با این مضمون که برچسب‌های گره‌ها بین همسایه‌ها به اشتراک گذاشته می‌شود.



# پروژه‌های ترکیبی و استفاده از داده‌های مختلف برای افزایش دقت و کارایی

[https://link.springer.com/chapter/10.1007/978-981-15-1925-3\\_20](https://link.springer.com/chapter/10.1007/978-981-15-1925-3_20)

استفاده‌ی همزمان از داده‌های مختلف در شبکه‌های اجتماعی مانند متن‌ها و تصاویر منتشر شده، ارتباطات موجود و سایر اطلاعات درون شبکه این امکان را برای ما فراهم می‌کند تا بتوانیم با دقت بیشتری ویژگی‌های شخصیتی و فردی افراد را استخراج کنیم.