

Analiza Rețelelor Media Sociale - Instagram Hashtag Analysis

Bucătaru Andreea A2

Extragerea informațiilor

Instagram-ul, la fel ca alte rețele sociale, folosește hashtag-urile ca o metodă de a promova idei și de a filtra postările. Am ales să fac analiza hashtag-urilor din domeniul artei, folosind o listă de hashtag-uri cheie: `'art', 'blackandwhite', 'artwork', 'artistsoninstagram', 'artexhibition', 'artphotography', 'artlovers'`.

Putem găsi ușor postările cu un anumit hashtag accesând link-ul `https://www.instagram.com/explore/tags/the_tag_here/?__a=1`, care va returna un JSON cu informațiile despre postări. Folosesc funcția `snowball` pentru a accesa link-urile și a extrage informațiile pe care apoi le parsez cu funcția `json2posts`.

The screenshot shows a Jupyter Notebook interface with two main panes. The left pane displays the Python code in a file named `insta.py`. The code imports various libraries and defines functions for requesting data from Instagram, performing a snowball search, and parsing JSON results. The right pane shows the output of the code execution. It lists the hashtags analyzed and their respective post counts: art (2627), blackandwhite (2630), artwork (2554), artistsoninstagram (2683), artexhibition (2372), artphotography (2505), and artlovers (2662). Below this, it provides a sample of nodes and edges, and a detailed table of edges grouped by weight, source, and target.

| weight | source | target |
|--------|---------------|------------------------|
| 17075 | artist | ipaddrawing |
| 87234 | sony | tvseries |
| 13878 | artexhibition | emergingartistplatform |
| 28852 | atxmodel | bodysuitoutfit |
| 11541 | artconsultant | banksy |
| 7149 | art | artwork |
| 7054 | 204 | artist |
| 7075 | 167 | artistsoninstagram |
| 7731 | 167 | drawing |
| 16454 | 163 | artist |
| 8892 | 137 | painting |
| 27034 | 130 | artwork |
| 19250 | 119 | artistsoninstagram |
| 16769 | 113 | artist |
| 9360 | 111 | art |
| 204 | 1 | sketch |

Figure 1: Extragerea datelor din Instagram

Crearea grafurilor

Am folosit librăriile *pandas* și *networkx* pentru a crea, procesa și exporta grafurile. Hashtag-urile reprezintă nodurile. Două noduri vor fi conectate printr-o muchie dacă au fost postează împreună măcar odată. Ponderea (weight) unei muchii este frecvența cu care o pereche de hashtag-uri au apărut în postări.

Funcția *write_graphml* din librăria *networkx* crează un fisier de tip *GraphML* cu nodurile și muchiile create anterior, care poate fi importat în Gephi:

The screenshot shows the Gephi 0.9.2 - Project 1 interface. The top menu bar includes File, Workspace, View, Tools, Window, and Help. Below the menu is a toolbar with three tabs: Overview, Data Laboratory (selected), and Preview. A sub-menu for the Data Laboratory tab shows 'Initial_graph' (selected). The main workspace is a Data Table view with the following columns: Id, Label, Interval, and weight. The table contains approximately 40 rows of data, each representing a hashtag. The 'Configuration' tab is also visible at the top of the table view. At the bottom of the table view, there are several buttons: Add column, Merge columns, Delete column, Clear column, Copy data to other column, and Fill column with a value.

| Id | Label | Interval | weight |
|-------------------------|-------------------------|----------|--------|
| art | art | | 425 |
| artist | artist | | 216 |
| artistsofinstagram | artistsofinstagram | | 16 |
| artistsoninstagram | artistsoninstagram | | 187 |
| artoftheday | artoftheday | | 109 |
| cartoon | cartoon | | 8 |
| chameleon | chameleon | | 1 |
| colour | colour | | 6 |
| copic | copic | | 3 |
| creative | creative | | 65 |
| doodle | doodle | | 22 |
| doodler | doodler | | 1 |
| doodles | doodles | | 6 |
| doodlesofinstagram | doodlesofinstagram | | 2 |
| doodling | doodling | | 2 |
| drawing | drawing | | 175 |
| drawwithjazza | drawwithjazza | | 1 |
| funny | funny | | 5 |
| illustration | illustration | | 108 |
| illustratorsoninstagram | illustratorsoninstagram | | 6 |
| instaart | instaart | | 100 |
| lineart | lineart | | 8 |
| sharpie | sharpie | | 1 |
| shrimpy | shrimpy | | 1 |
| shrimpydoodles | shrimpydoodles | | 1 |
| sweiperart | sweiperart | | 1 |
| vexx | vexx | | 1 |
| zhc | zhc | | 1 |
| 1312 | 1312 | | 1 |
| aerosmith | aerosmith | | 1 |
| caricatures | caricatures | | 2 |

Figure 2: Baza de date importată în Gephi

Am aplicat culori în funcție de ponderi și labels pentru a observa mai ușor clusterele și cele mai frecvente noduri.

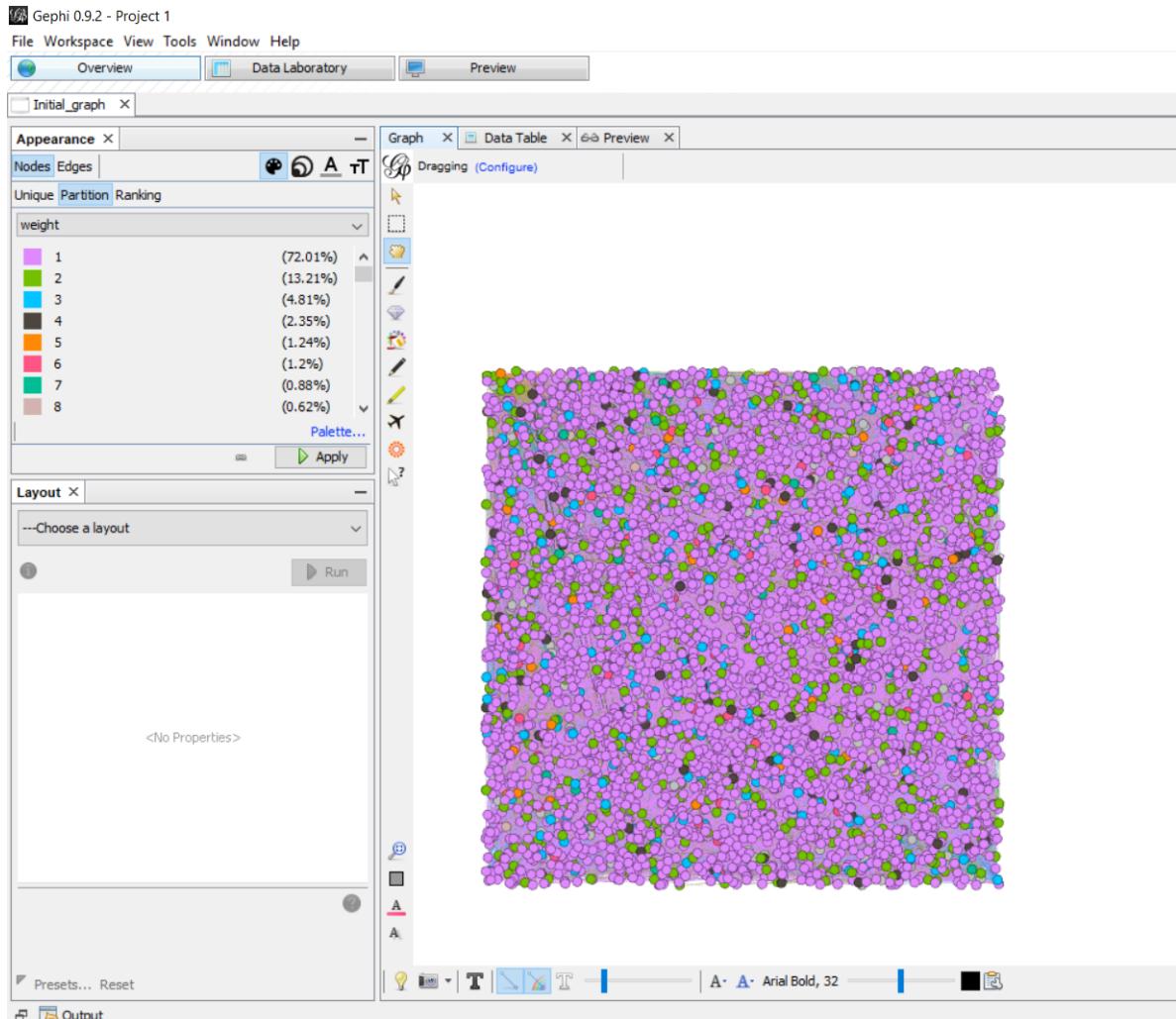


Figure 3: Graful inițial

Aplicarea algoritmilor

Am aplicat algoritmii *ForceAtlas* și *FruchtermanReingold*, care grupează în clustere nodurile strâns legate între ele (având ponderile mari). Prin acești algoritmi se pot observa foarte ușor legăturile între hashtag-uri și se pot analiza informațiile primite de la Instagram.

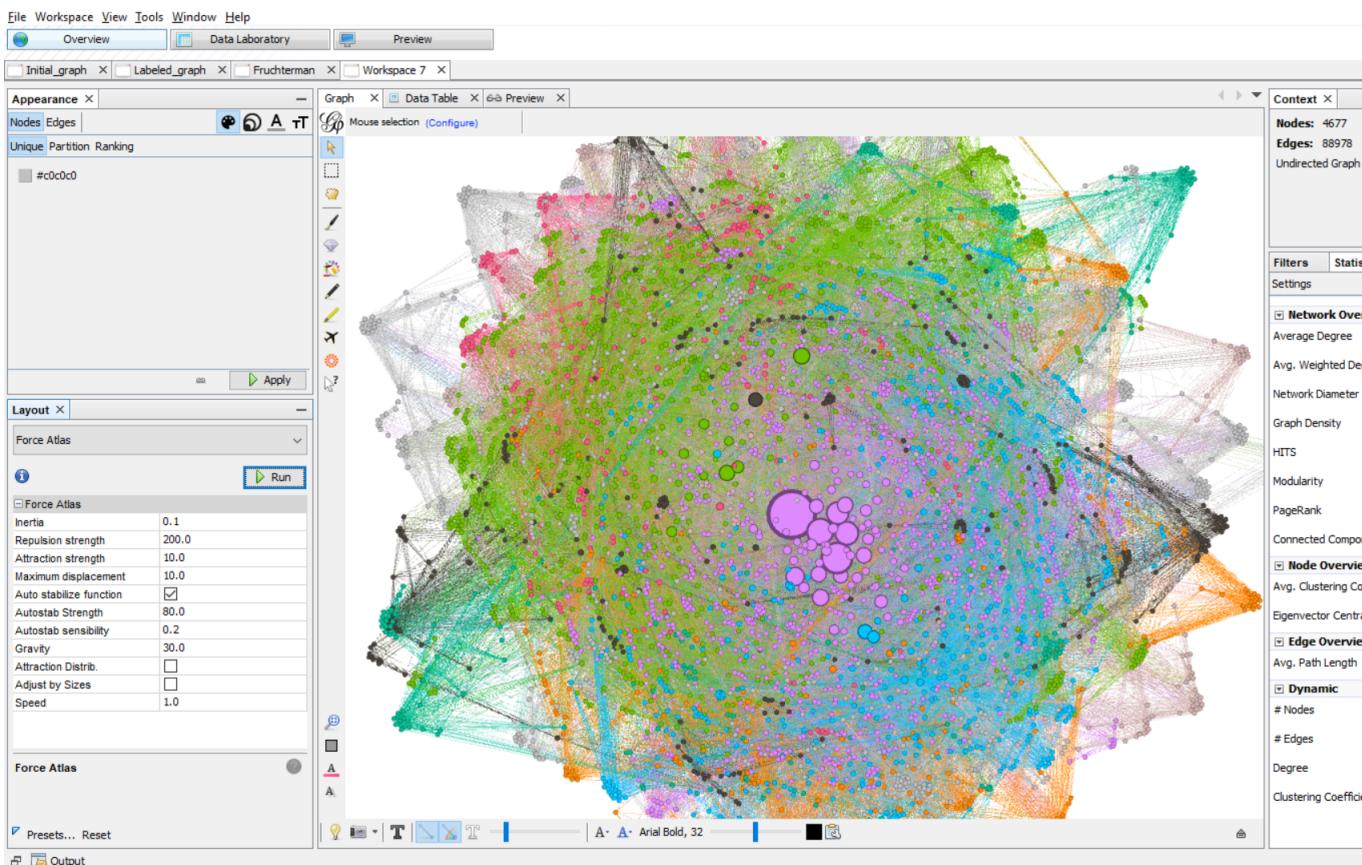


Figure 4: Force Atlas layout

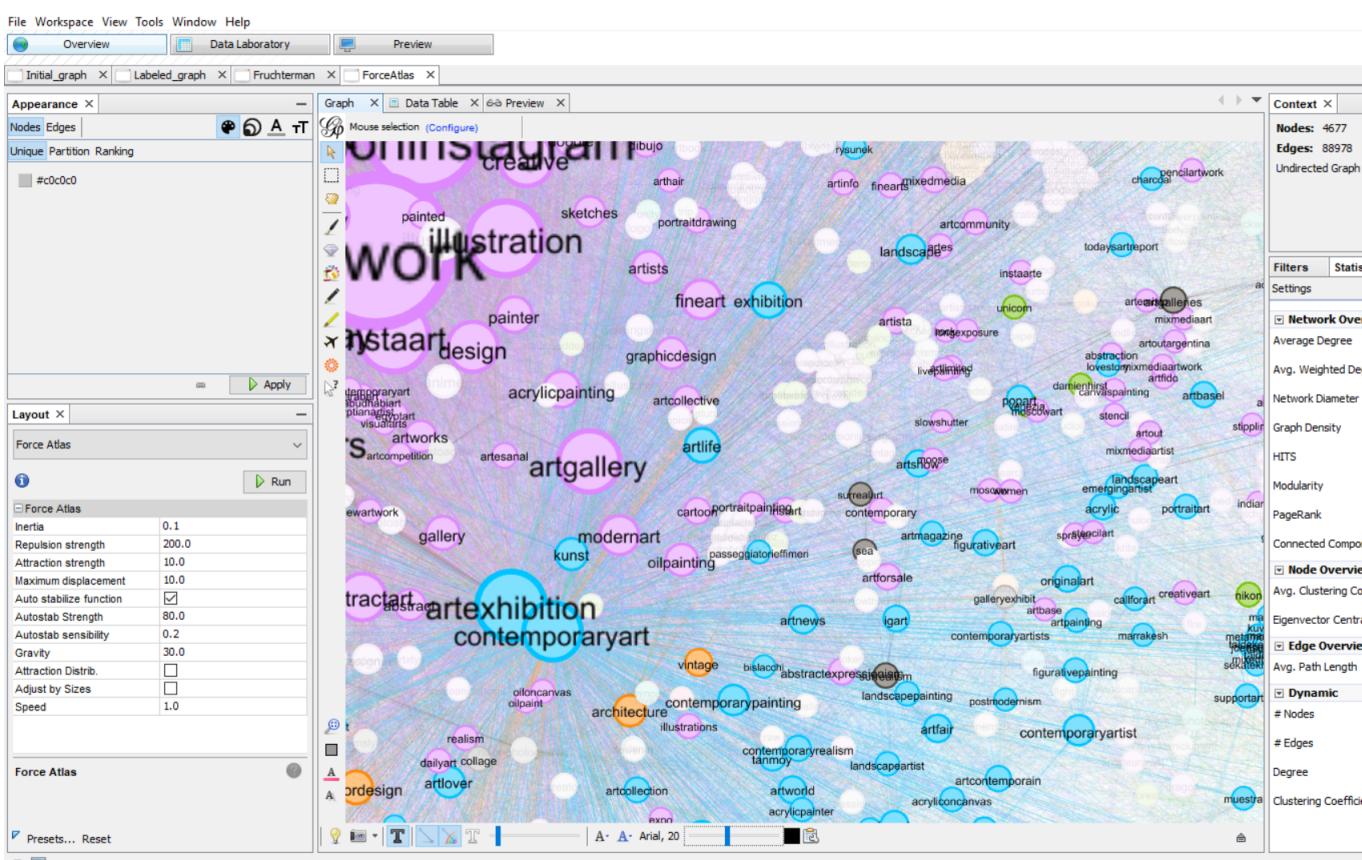


Figure 5: Zoom pe Force Atlas layout

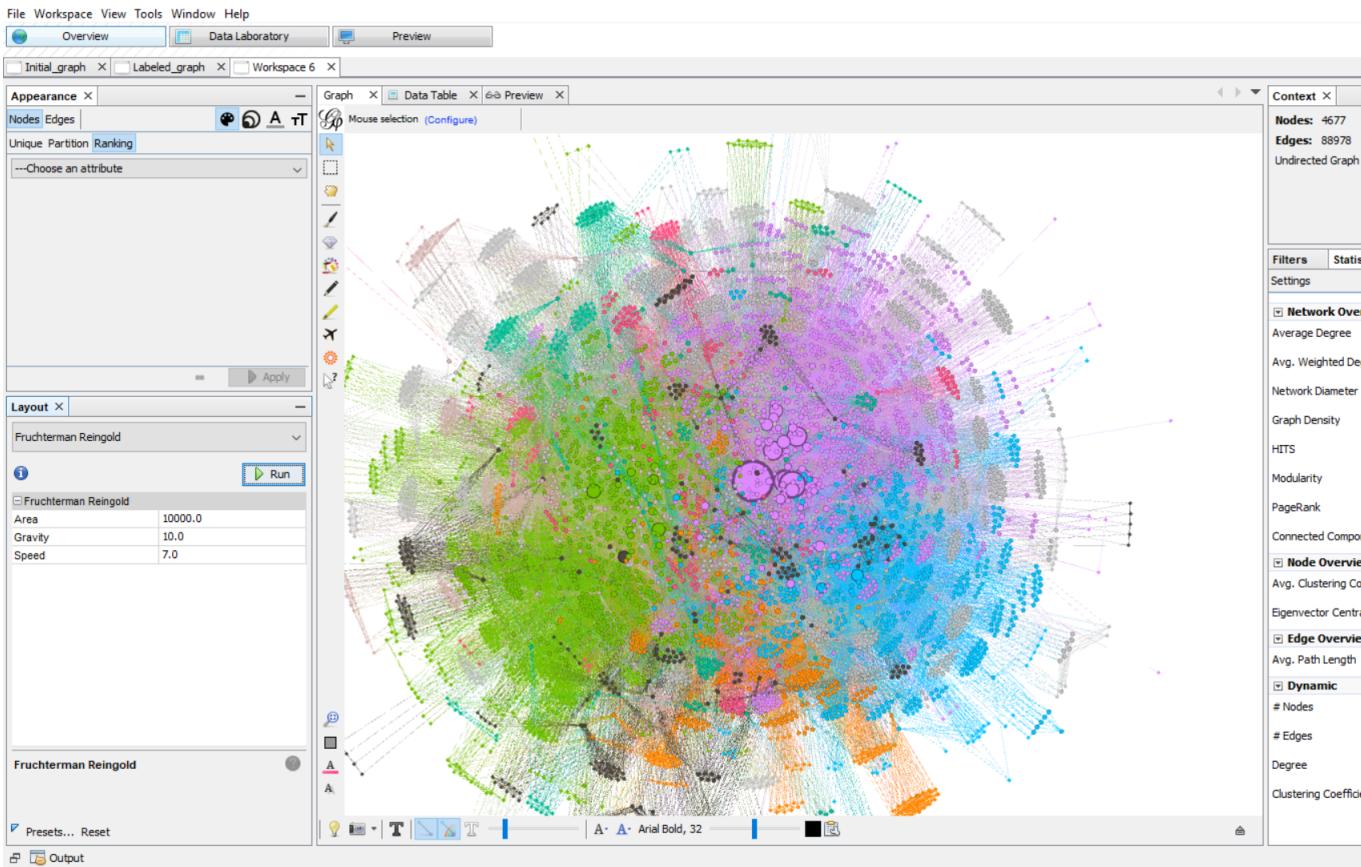


Figure 6: Fruchterman Reingold layout

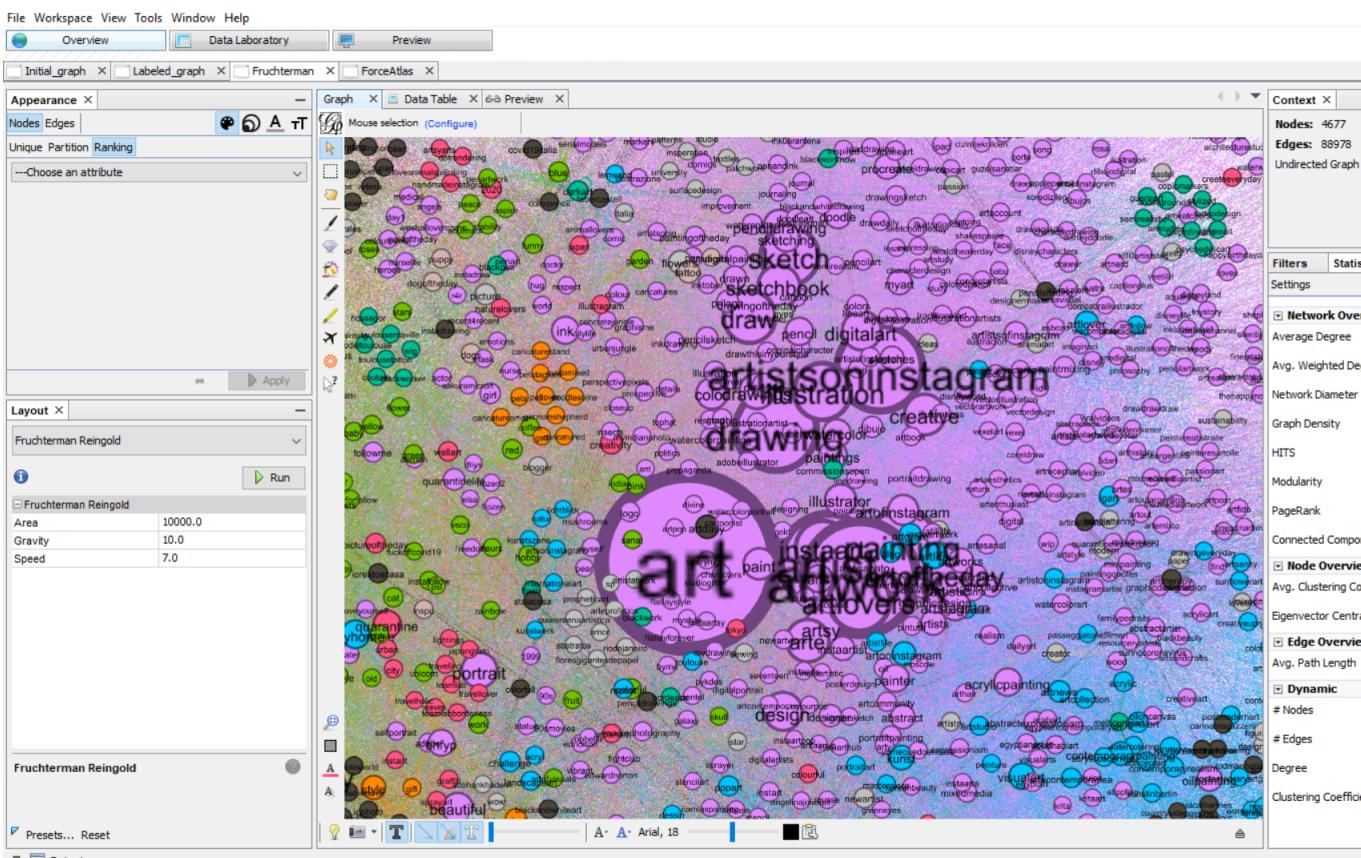


Figure 7: Fruchterman Reingold layout