

Laboratorio di Web Scraping
CRAFTY:
Clustering and scRAping For biTcoin deanonymization
Progetto di Fine Corso A.A. 2022/23
Notes

In questo documento verranno raccolte alcune note di chiarimento relative al progetto assegnato:

1. Nel progetto si chiede di considerare i 10 cluster di dimensione maggiore ottenuti dalla clusterizzazione e di prendere in considerazione gli indirizzi contenuti in questi cluster e provare a deanonimizzarli, considerando alcuni siti web. Il processo di deanonimizzazione deve essere effettuato utilizzando una strategia “best effort”, ovvero è possibile che la deanonimizzazione di alcuni cluster non sia possibile a causa dell’elevato tempo di scraping/ a causa di meccanismi anti-scraping come l’uso di captcha da parte dei server (non trattati a lezione). Comunque abbiamo verificato come sia possibile deanonimizzare almeno la metà dei cluster ottenuti.
2. Una volta deanonimizzati i cluster, si richiede di descrivere nella relazione quali servizi sono stati associati ai vari cluster. In particolare, se si ottengono servizi diversi utilizzando i due siti proposti (<https://www.wallexplorer.com/> e <https://bitinfocharts.com/>), provare a motivare perché questo accade.