

1. 需求数据量：5 万条推文+每个推文作者的 following/friend 数据
若无法获取 friend 数据，则使用 follower 数据替代
2. 数据要求：
 1. 每个用户最多保留 20 条推文
 2. 推文必须包含 5 中的关键字（至少一个）
 3. 推文的时间在 2020.11.3-2021.1.6 之间，优先保留靠近 2021.1.6 的推文
 4. 若无满足 2.2 要求的推文数据，则保存距离 2021.1.6 最近的 3 条推文
 5. following 数据来自于数据收集过程中
 6. 每个用户的 friends 数据至多不超过 100
3. 数据搜索方式。以 seed user 为起始节点，以 friend/following 作为边，进行广度优先搜索(BFS)。每个用户节点最多选取 200 个关系(friend+following)作为边。

Input: initial seed user u_0 in a user cluster

Output: user information set F

$u_0.layer \leftarrow 0$; // designate seed user as layer 0

$S \leftarrow \{u_0\}$; // set of users to expand

$u_0.expanded \leftarrow False$;

$F \leftarrow \emptyset$;

while $S \neq \emptyset$ **do**

$u \leftarrow S.pop()$; // expand with user u

$T(u) \leftarrow get_tweet(u)$;

$P(u) \leftarrow get_property(u)$;

if $u.layer \geq 3$ **or** $u.expanded == True$ **then**

$F \leftarrow F \cup u(T, P, N = \emptyset)$;

continue; // three layers max

$u.expanded \leftarrow True$;

$N^f(u) \leftarrow get_friend(u)$;

$N^t(u) \leftarrow get_follower(u)$;

$N(u) \leftarrow \{N^f(u), N^t(u)\}$;

$F \leftarrow F \cup u(T, P, N)$;

$S \leftarrow S \cup N^f(u) \cup N^t(u)$;

for $y \in N^f(u) \cup N^t(u)$ **do**

$y.expanded \leftarrow False$;

$y.layer \leftarrow u.layer + 1$;

Return F ; // obtained one cluster of user information

4. Seed User {显示姓名，账户名，个人主页 url}
 - Tucker Carlson, @TuckerCarlson, <https://twitter.com/TuckerCarlson>
 - Michael Moore, @MMFlint, <https://twitter.com/MMFlint>
 - Nina Turner, @ninaturner, <https://twitter.com/ninaturner>

Sarah Silverman,@SarahKSilverman,https://twitter.com/SarahKSilverman
Candace Owens,@RealCandaceO,https://twitter.com/RealCandaceO
Barack Obama,@BarackObama,https://twitter.com/BarackObama
Kayleigh McEnany,@kayleighbcenany,https://twitter.com/kayleighbcenany
Kamala Harris,@KamalaHarris,https://twitter.com/KamalaHarris
Lindsey Graham ,@LindseyGrahamSC,https://twitter.com/LindseyGrahamSC
Ilhan Omar,@IlhanMN,https://twitter.com/IlhanMN
Charlie Kirk,@charliekirk11,https://twitter.com/charliekirk11
Sarah Huckabee Sanders,@SarahHuckabee,https://twitter.com/SarahHuckabee
Official Rush Limbaugh,@OfficialRushUSA,https://twitter.com/OfficialRushUSA
Bernie Sanders,@BernieSanders,https://twitter.com/BernieSanders
Elizabeth Warren,@SenWarren,https://twitter.com/SenWarren
LeBron James,@KingJames,https://twitter.com/KingJames
Michelle Obama,@MichelleObama,https://twitter.com/MichelleObama
Danny DeVito,@DannyDeVito,https://twitter.com/DannyDeVito
Lady Gaga,@ladygaga,https://twitter.com/ladygaga
Sean Hannity,@seanhannity,https://twitter.com/seanhannity
Mark Ruffalo,@MarkRuffalo,https://twitter.com/MarkRuffalo
Alexandria Ocasio-Cortez,@AOC,https://twitter.com/AOC
John Legend,@johnlegend,https://twitter.com/johnlegend

5. 关键词列表:

Trump2020LandSlide
TrumpCrimeFamily
DonaldTrump
Republican
BidenForPresident
SleepyJoe
JoeBiden
Democrats
BernieWon
NeverBernie
BernieSanders
Sanders
Election2020
SocialistSanders
BernieLosesAgain
SandersForPresident
MAGA
KeepAmericaGreat
AmericaFirst
TrumpTrain
NeverTrump
DumpTrump

TrumpResign
NotMyPresident
BidenHarris2020
VoteBlue
JoeAndKamala
BidenUnity
NeverBiden
BidenCrimeFamily
NotMyPresidentBiden
BidenLied
FeelTheBern
Bernie2020
NotMeUs