הסבר כיצד בחרנו את המשפיענים

תחילה, חשבנו להשתמש באלגוריתם "climbing hill" כפי שראינו בהרצאה. כאשר עשינו זאת, הבנו שהוספת האילוץ של התקציב גרמה לכך שהבעיות לא לחלוטין שקולות והבנו שנצטרך להתאים את האלגוריתם לבעיה שקיבלנו.
לכן, שינינו אותו כך שימשיך להוסיף מועמדים (לפי השפעתם השולית הגבוהה ביותר) עד לתקציב המבוקש אך ראינו שנוצרות מספר בעיות אחרות ביניהן: חוסר ניצול של כל התקציב, תוצאות נמוכות מהרף שביקשו וזמן ריצה מאוד ארוך.

בשלב זה חשבנו לנסות למצוא תת קבוצה התחלתית ("select_influencers") של מועמדים על פי קריטריונים מסויימים שיתנו להם עדיפות, שמהם נבחר את הצמתים שיהוו את הקבוצה הסופית ("climbing hill").

על מנת לבחור קריטריונים אלו, עברנו על ההרצאות וקראנו מאמרים באינטרנט העוסקים בבעיות דומות.
בסופו של דבר, לאחר מספר נסיונות של הרצות שונות וויזואליזציות על הגרף מצאנו כי הבחירה הטובה ביותר היא שימוש במדדי BETWEENESS,DEGREE כלומר, בחרנו את הצמתים שהם הכי מרכזיים מהבחינה שהכי הרבה מסלולים קצרים ביותר בגרף עוברים דרכם ומתוכם בחרנו את אלו שמספר השכנים שלהם הוא הגדול ביותר (תחילה בחרנו את 400 הצמתים עם BETWEENESS הכי גבוה ואז 200 בעלי הדרגה הכי גבוהה -המספר נקבע לאחר סימולציות רבות).
במהלך העבודה, הייתה מחשבה להכליל בקריטריונים גם את קריטריון הOVERLAP ובכל שלב לבחור לתת-הקבוצה הראשונית גם צמתים שקריטריון האוברלאפ נמוך ביניהם. כאשר ביצענו זאת, לא ראינו הבדלים גדולים בין התוצאות (גם בויזואליזציות של בחירת הצמתים) – ראינו שבכל מקרה נבחרים צמתים שיחסית רחוקים זה מזה ולכן ויתרנו על כך (הפונקציה עדיין מופיעה בפתרון שלנו).
מויזואליזציות שערכנו הבחירה הזו הביאה לנו קבוצה של צמתים שבעיקר "מגשרים" בין אשכולות בגרף וגם צמתים שיחסית מרכזים בתוך האשכולות. להערכתנו הצמתים האלו יהיו הנגישים ביותר לצמתים אחרים ולחלקים שונים בגרף ויוכלו לגרום לפיזור מיטבי של ההשפעה בגרף.

כדי להתמודד עם בעית ניצול התקציב החלטנו להשתמש ברעיון של בעיה מוכרת – "בעית התרמיל".
אומנם פונקצית המטרה שלנו במקרה זה אינה לינארית, אבל חשבנו שבעקבות מאפיינים דומים בין הבעיות, במקום לנסות לבחור את המועמד שממקסם את תועלת ההשפעה השולית נוכל למקסם את היחס בין תועלת ההשפעה השולית של צומת לבין מחירו. בצורה זו נוכל לנצל את התקציב שלנו בצורה נכונה יותר כי בעית האופטימיזציה תשאף לבחור מועמדים בעלי השפעה גבוהה ועלות נמוכה.

בנוסף, דאגנו לכך שהפונקציה שלנו שאותה אנו ממקסמים (פונקצית היחס לעיל) אכן סאב-מודולרית כנדרש לפעולה מיטבית של אלגוריתם "climbing hill" (אנו לוקחים בחשבון שבעקבות אילוץ התקציב הבעיות אינן שקולות אבל קיווינו שתכונה זו עשויה לעזור לאלגוריתם למצוא פתרון טוב יותר).

בשלב הבא רצינו לוודא שנקבל את האומדן המדויק ככל האפשר לתועלת ההשפעה השולית של כל מועמד ולכן הציון שלו התבסס על ממוצע של 100 סימולציות. מספר זה נקבע לאחר מספר נסיונות שבהם ניסינו לשלב בין ייעול זמן ריצה לבין דיוק. כידוע ממוצע הוא אומד עקיב וחסר הטייה לתוחלת ההשפעה ולכן בחרנו להשתמש בו.
במהלך כל סימולציה, בשלב השישי של ההדבקה בחרנו לסכום את ההסתברויות להדבקה של כל צומת ולא לסמלץ אותן כפי שנהגנו בחמשת השלבים הראשונים. זאת משום שתוחלת ההשפעה שווה לסכום ההסתברויות להדבקה של צמתים נוספים ולכן מספיק היה לחשב אותן ולא ראינו צורך בסימלוץ של שלב הדבקה נוסף (לא ראינו הבדלים מאוד משמעותים בדרך זו לבין לעשות סימולציה לשלב ה6). בסופו של דבר בחרנו את הקבוצה שהניבה לנו את התוצאה הגבוהה ביותר באלפי סימולציות שערכנו באמצעות פונקציה שנקראת "simulate_influence_scoring".