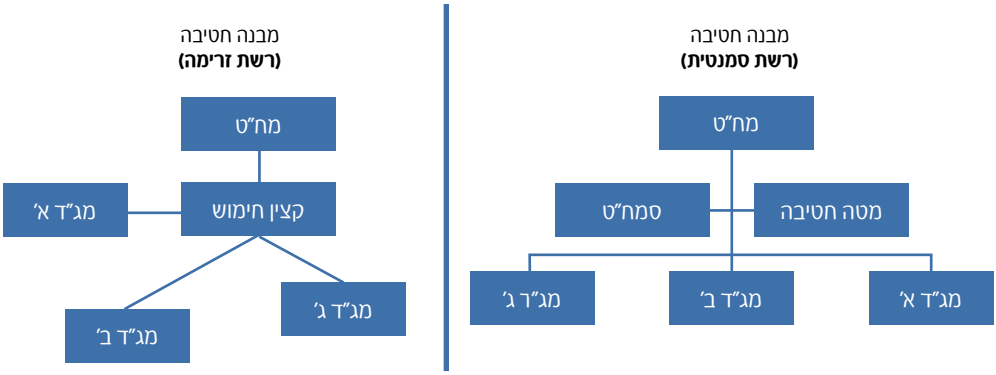
ניתוח מודיעין רשתי בעידן הביג דאטה

רס"ן א' - משרת באמ"ן

מהי רשת ומה תרומתה למחקר?

מחקר רשתי מאפשר, באמצעות התבוננות על התנהלות של רשתות, הבנה מעמיקה יותר של תפקודם של מושאי המחקר ושל התופעות הנצפות ברשת תוך זיהוי "מרכזי כובד" (גורמים מרכזיים משפיעים).

איור 8: רשת סמנטית לעומת רשת זרימה



המחשה: לכאורה, הגורם החשוב בחטיבה הוא המח"ט (רשת סמנטית) אך בפועל, כשנגמרת התחמושת, ניתן לראות שברשת הזרימה (הקשרים שנוצרים בין הגורמים בחטיבה) קצין החימוש הוא מרכז הכובד.

מאמר זה עוסק ברשתות. לצורך השפה המשותפת, *"רשת"* (Network) היא כל מידע שמורכב מ*"צמתים"* ("nodes") ו*"קשתות"* ("edges"). הצומת הוא נקודה ברשת והקשת היא הקו המחבר בינה לבין נקודות האחרות.

מקרה קלסי הוא רשתות חברתיות (פייסבוק, טוויטר וכדומה), טלפוניה וכולי, אך גם קשרים בין שרתי מחשבים, מסירות במשחק כדורגל ואינטראקציה בין חלבונים אפשר להציג כרשת. מחקר ארגוני הוא כמעט בהכרח מחקר רשתי, שכן ניתן לאפיין את הארגון לפי האינטראקציות המתרחשות בו. הרשת היא בעצם כלי ביטוי המאפשר להגדיר את מהות קשרי הגומלין ואופיים של היחסים שבין מרכיבי המערכת.

במה שונה הניתוח הרשתי (SNA) מהניתוחים שנעשו עד היום במודיעין? בהיעדר תפיסה או יכולת של ביג דאטה, עיקר העיסוק המחקרי היה ב־link analysis. כלומר, מחקר קשרים, שהוא מחקר נקודתי של עוגנים והצמתים שמסביבם. לעומת זאת, גישת ה־SNA מבקשת להסתכל באופן רחב ככל שניתן על הרשת ועל סמך אלגוריתמיקה לזהות את מרכזי הכובד (data-driven)

ולא על סמך אינטואיציות חוקר. גישה כזו מאפשרת "מבט מרחוק" על מושא המחקר והרחבתו במקרה הצורך של מושא המחקר (קהילות של אנשים ולא אנשים או ארגונים בודדים). הניתוח הרשתי גם מאפשר להגיע לתובנות שאינן מובנות אפילו למושא המחקר עצמו. לא תמיד אנו מודעים לצווארי הבקבוק בתהליכים שאנו מקיימים, הפורמליים והלא פורמליים, ואיננו מודעים לזהות מרכזי הכובד בארגון שבו אנו חברים בכל רגע נתון.

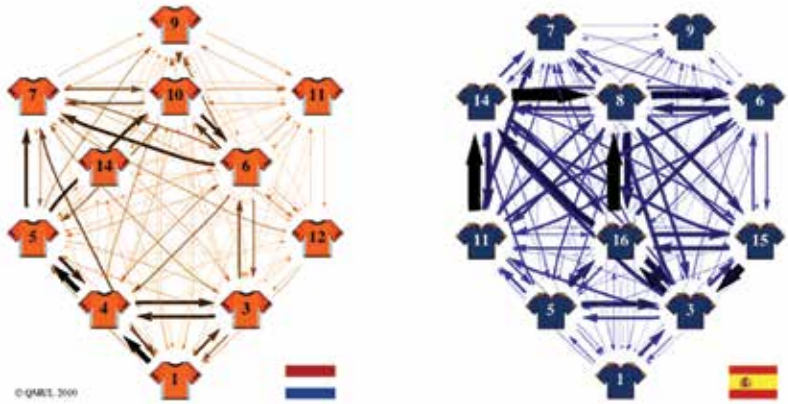
המחקר הרשתי מסתכל באופן רחב על הרשת כדי לזהות מרכזי כובד בהתבסס על דאטה ולא על סמך אינטואיציה של חוקר.

הניתוח הזה מאפשר להגיע להבנות שאינן ידועות למושא המחקר עצמו.

עידן הביג דאטה גרם לקפיצת מדרגה משמעותית במחקר הרשתי. מחד גיסא, אִתגר היצף המידע את יכולות החוקר להבין הקשרים ולראות תמונה רחבה בכלים המסורתיים, ומאידך גיסא יכולות המחשוב והאלגוריתמיקות שפותחו מאפשרות לנתח כמויות אדירות של מידע ולספק תובנות מועילות. הבניית המידע כרשת מאפשרת לנו לייצר תובנות שמגבלותינו האנושיות מונעות מאיתנו. הדמיון האנושי מוגבל ומתקשה להכיל מערכת מורכבת. ניתוח המערכת כרשת מאפשר לקמ"ן "לספר סיפור על הדאטה". כלומר, לתאר תופעות בהקשר ובזמן אמת.

בניתוח המסורתי, הקמ"ן מניח תרחיש (scenario) ואז מחפש אותו בנתונים. לעומת זאת, ניתוח הרשת מאפשר יכולת הפוכה: סידור הדאטה כרשת והבנת המערכת שלפיה משחררים את החוקר מהצורך להניח הנחות לצורך הבנת הרשת. כך, ניתוח המערכת נצמד לנתונים המאפשרים להפריך או לאשש בקלות יחסית את תמונת המודיעין, וזאת בניגוד לתזה הנשענת על אינטואיציה.

איור 9: "ניתוח רשת" של משחק כדורגל



ניתוח רשתי של משחק כדורגל ספרד-הולנד באמצעות תיעוד המסירות (עובי הקו הוא כמות המסירות). ניתן להצביע על השחקנים המרכזיים (מרכזי כובד), איזה צד (ימין/שמאל) היה דומיננטי יותר (או בשפה הצה"לית, מה היו הדפ"אות ומהו הדפ"ן).

רכזות וחוק ה-20/80

רשתות גדולות מייצגות מערכות מורכבות מאוד אך אין זה אומר שאין הן פועלות לפי חוקים. עם התפתחות המחקר בתורת הרשתות התגלה כי עולם הרשתות אינו אקראי כפי שהיה נהוג לחשוב והבנת חוקים אלו מאפשרת מחקר אפקטיבי.

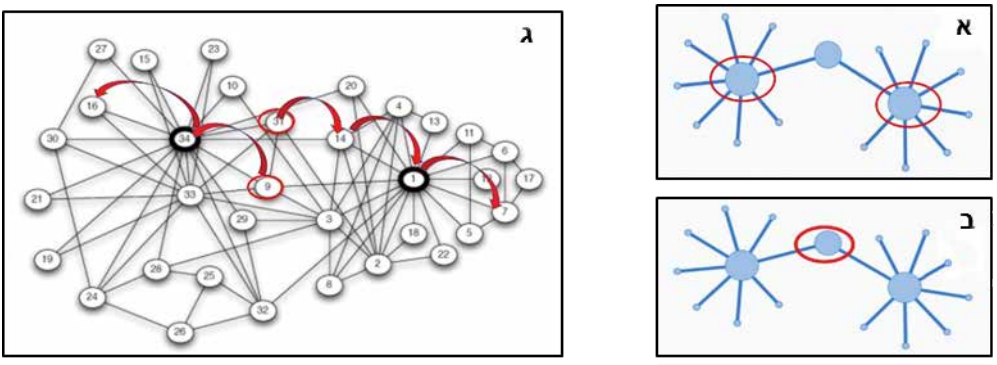
חוק 1: powerlaw

נראה שמעט צמתים "פופולריים" מאוד "שולטים ברשת כולה".3 דוגמאות לצמתים כאלו באינטרנט הם אתרים כמו Google. ברשת של אנשים, מדובר באנשים שמכירים את כולם וכולם מכירים אותם. המחברים הללו מכונים "רכזות". הרכזות קיימות כמעט בכל רשת בעולם, כאשר מספר נמוך של אנשים שאחראים לאחוז גבוה של שיחות נכנסות ויוצאות ועוד. למעשה, ניתן לומר שמתקיים כאן חוק 20/80 של פארטו (Pareto) לגבי התפלגות הקישורים: 80% מהקישורים שייכים ל-20% מהצמתים ברשת (לרכזות). תופעה זו נקראת במחקר הרשתות Power Law או Zipf's Law. איתור צמתים אלה מאפשר לזהות את מרכזי הכובד של הרשת ועל ידי אחיזה בהם לאחוז כמעט בכל הרשת. כדי לאתר צמתים אלו, נדרש שימוש באלגוריתמיקת SNA.

Social network analysis - SNA

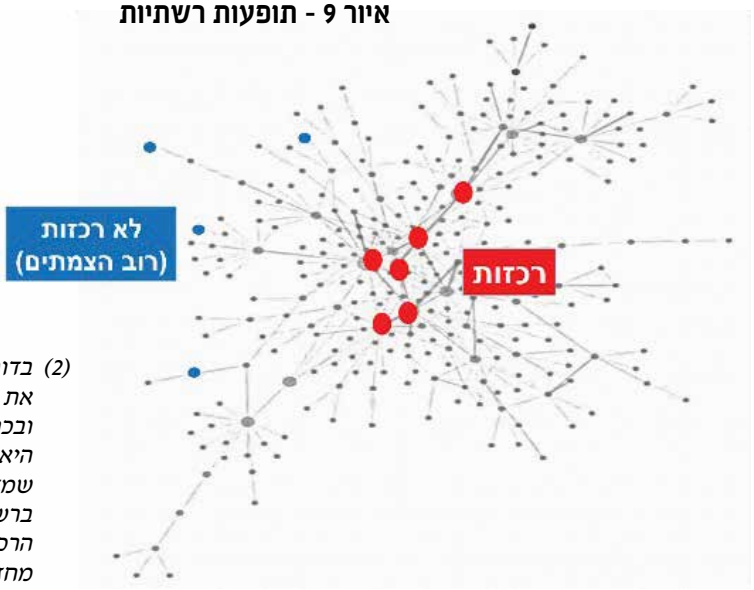
SNA הוא עולם האלגוריתמיקה המאפשר להבנות ולחקור את הרשת. ישנם עשרות, אם לא מאות, אלגוריתמים לזיהוי "רכזות" ברשת, אך הבסיסיים שבהם הנמצאים בשימוש רחב הם:

- Degree Centrality: מודד את כמות הקשרים של הצומת. ההיגיון הוא שהצומת מרכזי יותר ככל שהוא מקושר ליותר גורמים. דוגמה אפשרית בארגון: הגורם המחזיק את סדר היום בארגון (חמ"לים, לשכות).
- Betweenness Centrality: מודד את כמות המסלולים הקצרים ביותר ברשת העוברים דרך הצומת. ההיגיון הוא שלהיות גורם מגשר הופך את הצומת למרכזי. דוגמה אפשרית בארגון: הגורם המחבר אזורים מנותקים, מקרב את הפריפריה הארגונית פנימה, המקור לכניסה של רעיונות חדשים ויצירתיות.



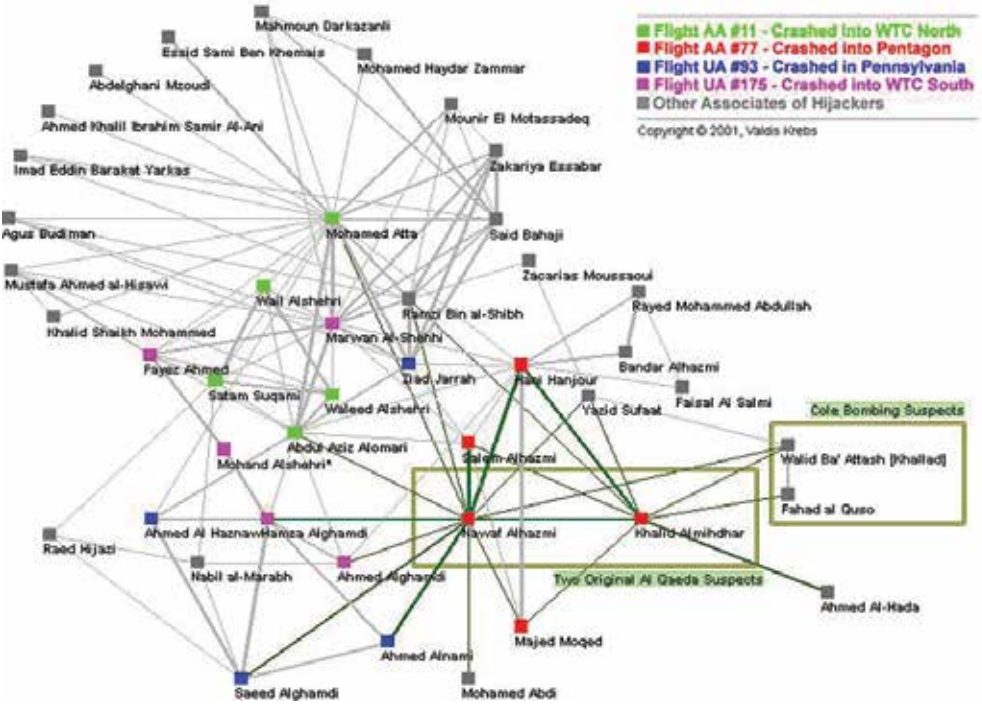
(1) בדוגמאות לעיל: העיגול האדום מציין את המובילים במדדי המרכזיות האלה (א) Degree, (ב) Betweenness, (ג) Closeness. החיצים האדומים משמשים להמחשת המרחק בין הצמתים המרכזיים (צומת 31 וצומת 9) לשאר הרשת.

איור 9 - תופעות רשתיות



(2) בדוגמה ניתן לראות באדום את הגורמים המרכזיים ברשת ובכחול את שאר הצמתים. היא ממחישה גם powerlaw שמדגים שלרוב הצמתים ברשת קשרים בודדים ואילו הרכזות (שמהוות מיעוט) מחזיקות את רוב הקשרים.

איור 10: מחקר רשתי של רשת הטרור אל־קאעידה בפיגועי 11.9.2001

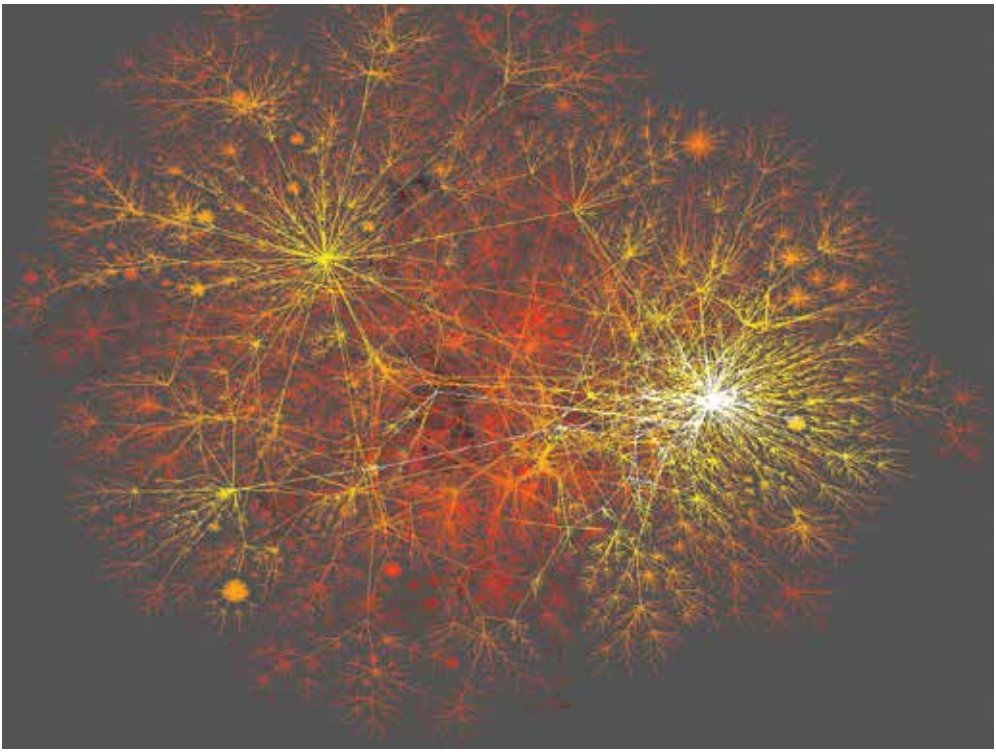


- Closeness Centrality: מודד את מרחק הצומת משאר הצמתים ברשת. ההיגיון הוא שצומת שנמצא ב"מרכז" הוא מרכזי גם אם איננו קשור לגורמים רבים. דוגמה אפשרית בארגון: המיקום הופך את השחקן לדומיננטי בזכות הקרבה היחסית שלו לשחקנים האחרים.

כדי לחקור ארגון באמצעות ניתוח רשתי, אין די בבדיקת הגורמים המרכזיים, שכן יכולות להיווצר הטיות. דוגמה לכך היא מחקר אמריקאי שנעשה על רשת הטוויטר במצרים כדי לזהות את הגורמים המובילים במהפכת "האביב הערבי" ב־2011. הגורמים המרכזיים ביותר היו ג'סטין ביבר וקייטי פרי, שתרומתם למוזיקת הפופ מוכחת, אך לאו דווקא להבנת גורמי הכוח מאחורי האירועים...

לשם זיהוי מרכז הכובד המעניינים אותנו, החוקרים, נידָרש לחוק 2 של הרשת: **רשתות מתקהלות** (מלשון: "קהילות").

הסתכלות קרובה יותר על המבנה הפנימי של הרשת מלמדת כי הרשת אינה אקראית אלא בנויה מצבירים. לצבירים אלה היגיון של הומופיליה (דומות, קרי הנטייה לחבור למי שדומה לך) או ריבוי אינטראקציות פנימיות. חלוקה שכזו של הארגון לקהילות יכולה להמחיש את מבנה הארגון "האמיתי", נכון לזמן הבדיקה.

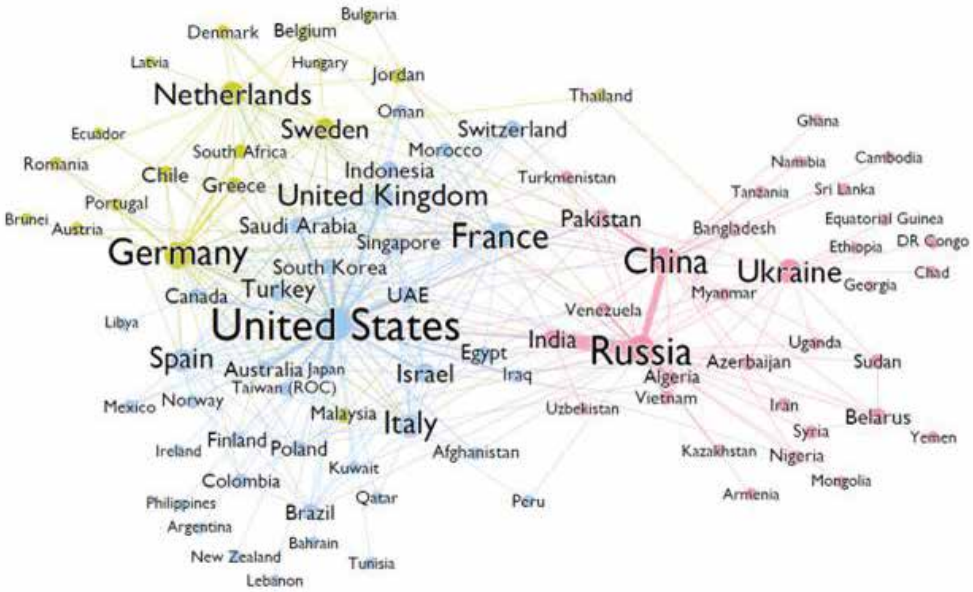


המחשה של רשת האינטרנט שבה ניתן להבחין בהתפצלויות מליבת הרשת שיוצרות ענפים ותתי־ענפים. כל ענף שכזה הוא קהילה המכילה תת־קהילות.

מיפוי הקהילות ומציאת גורמי המפתח בקהילות העניין מאפשרים למצוא את "מרכזי הכובד" הארגוניים. בדוגמה של כיכר תחריר, הגורמים המרכזיים שהניעו את המהפכה סביר שיימצאו בקהילה שאותה היינו מתייגים כ"קהילת תחריר" וסביר שזמרי הפופ היו בקהילה אחרת אותה היינו מתייגים כ"חובבי מוזיקה".

גם בתחום זה יש עשרות, אם לא מאות, אלגוריתמים עם הגיונות שונים לחלוקה לקהילות. אין כיום הגדרה "מדעית" לשאלה "מה היא קהילה?" ההגדרה הרווחת היא שקהילה היא קבוצת צמתים ברשת שכמות הקשתות שביניהם צפופה יותר מכמות הקשתות בצמתים אחרים ברשת.

איור 11 - מחקר רשתי של עסקאות נשק



מחקר SNA שנעשה על עסקאות נשק. ההמחשה מאפשרת לזהות את קהילות הסחר הקיימות בתחום ומיהם מרכזי הכובד בכל קהילה. במחקר זה, למשל, ניתן להבחין בשלוש קהילות מרכזיות: המזרחית (רוסיה/סין), האירופאית (גרמניה, אנגליה) והקהילה האמריקאית, אליה משתייכת ישראל.

מחקר דינמי

מחקר SNA ניתן לבצע על תקופת זמן מוגדרת ("snapshot"), שזוהי כיום השיטה הנפוצה, אך בשנים האחרונות חלה התפתחות במחקר הרשת והעיסוק בניתוח דינמי של הרשת לאורך זמן, גובר. תחום זה אף מהווה בסיס לגילוי שינויים (או בשפת המודיעין: "התרעות/התראות").

מהם הכלים הרלוונטיים לטובת גילוי שינויים?

מחקר רשת מעמיק מצביע שברשתות רבות אין יציבות ברמת הפרט. צמתים מופיעים ונעלמים וקשה לבסס עליהם שגרה. לעומת זאת, קהילות ברשת נוטות להיות יציבות יותר (גם אם הצמתים

שבתוכן נוטים להשתנות). כך, למשל, יהיה קשה לאפיין שגרה של מפקד בצורה אפקטיבית, אבל "קהילת הגדוד" היא יחסית יציבה . נוסף על כך, כל שינוי משמעותי בארגון/רשת סביר שיעורר חריגה של יותר מפרט בודד אחד. כלומר, זיהוי חריגה של פרט, אינו מעיד בהכרח על אירוע. אבל שינוי בהתנהלות הקהילה הוא בבחינת אירוע ("סימן מעיד") המחייב התייחסות.

לאור זאת, יתרונות ה־SNA לצורך זיהוי חריגה משגרה בפרט ובניתוח מודיעיני בכלל, הם:

- יכולת לניתוח אזורי המקרו (קהילה, רשת) ולא רק אזורי המיקרו (הפרט הבודד, צומת).
- הסתכלות "דרך עיני הדאטה" ללא צורך לעסוק בהשערות לגבי תרחישים, שבאופן טבעי, מוטות או מוגבלות על ידי הדמיון האנושי. נקודת המוצא היא שהעבר לא בהכרח מספר לנו את העתיד. לפיכך:
  - בניגוד לשיטות אחרות, לא נדרשות הנחות מוקדמות על הדאטה.
  - לא נדרש "אימון" (בוודאי לא משמעותי) של האלגוריתם על הדאטה.

לסיכום, יישום SNA (אלגוריתמיקת הרשת) מאפשר למודיעין לזהות מרכזי כובד וקהילות עניין. מהפכת הביג דאטה מאפשרת תשתית ואלגוריתמיקה לניתוח מיליוני נתונים בזמן קצר, כך שלא נדרש בהכרח לסמן עוגני עניין המבוססים על המחקר המסורתי ולהתחיל מהם את המחקר (Small Data בשיטה של תרחישים) אלא ניתן לקחת את כל הרשת, לנתחה בבת אחת ולאחוז בה לאורך זמן.

כך ניתן לגלות:

- מי הגורמים הדומיננטיים בזמן אמת (ללא צורך בידע מוקדם) לצורכי מטרות או מעקב.
- אילו גורמים חדשים מעניינים ולא מוכרים "צצו לפתע".
- כיצד פועל הארגון היריב "באמת" (לא לפי עץ המבנה הרשמי אלא לפי האינטראקציה ברשת) ולזהות דפ"א מתהווה, תמ"א ועוד.