

אלגוריתמים ברשתות - פרויקט גמר

אלעד ארמוני 206157323 אופק בג'רנו 318435195

[מקור הנתונים](#) [לינק לפרויקט ב-GITHUB](#)

נושא הפרויקט

השפעת מגפת הקורונה על ביצועי השחקנים ותוצאות המשחקים בטורנירי טניס ATP.

מבוא ומוטיבציה

טורנירי ATP מהווים את לב עולם הטניס המקצועני, ומתקיימים במגרשים שונים ברחבי העולם. מגיפת נגיף הקורונה (COVID-19), שהתפרצה בסוף שנת 2019, השפיעה בצורה משמעותית על תחומי החיים השונים, כולל עולם הספורט. עם פרוץ המגפה, בוטלו או נדחו טורנירי ATP רבים, ושחקני הטניס נאלצו להתמודד עם אתגרים חדשים כמו הגבלות על נסיעות, אימונים בתנאים מוגבלים והפסקת התחרויות באופן פתאומי. מגבלות אלה גרמו לשינויים משמעותיים בביצועי השחקנים ובדינמיקות של המשחקים. מחקרים שנערכו על השפעת מגפת הקורונה על תחומי הספורט השונים הראו כי האירועים הגלובליים הללו גרמו לירידה משמעותית בתחרויות ובביצועים של ספורטאים.

בעידן הנוכחי, הספורט המקצועני מושפע רבות מגורמים חיצוניים ומקרו-אירועים, כמו מגפת הקורונה. הבנת ההשפעות של אירועים כאלה על דינמיקות משחקים, ביצועי שחקנים ומבנה רשתות הספורט היא חיונית לא רק לצורך ניתוח היסטורי אלא גם לצורך קבלת החלטות עתידיות וטיפול במצבים דומים.

המטרה המרכזית של הפרויקט היא לנתח את השפעת מגפת הקורונה על טורנירי ATP ושחקני הטניס המשתתפים בהם, באמצעות כלים ואלגוריתמים מתחום רשתות וגרפים. המחקר יתמקד בזיהוי שינויים במדדי מרכזיות ואלגוריתמים שונים לזיהוי קהילות בתקופות שלפני ובזמן המגפה, כדי לזהות שינויים משמעותיים ולהבין את ההשפעות המלאות של המגפה על עולם הטניס המקצועני.

מאמרים בתחום

השפעת מגפת הקורונה על הספורט:

1. how coronavirus is affecting the media industry

- המאמר מתאר כיצד מגפת הקורונה השפיעה על אירועי ספורט, לרבות דחיות וביטולים של טורנירים מרכזיים כמו האולימפיאדה. השפעות אלו כללו גם את ההשפעה על השחקנים והקהל, ואת ההתמודדות של ארגוני ספורט עם השינויים. המאמר מצביע על כך שהמגפה גרמה לשיבושים רבים באירועי ספורט מרכזיים ולהפסקות פעילות זמניות, דבר שהשפיע על תפקוד הארגונים ועל ביצועי השחקנים.
- [World Economic Forum](#)

2. Sports as a service: Will COVID-19 change how we watch sports?

- מאמר נוסף מצביע על השפעות כלכליות קשות על תעשיית הספורט, כולל אובדן הכנסות משידורים, כרטיסים וחסויות. המאמר מסביר כיצד המגפה שינתה את דפוסי הצפייה בספורט ואת המודלים הכלכליים של התעשייה. השפעות אלו הובילו לשינויים משמעותיים בהתנהלות הארגונים וביכולת שלהם להתמודד עם המצב החדש.
- [World Economic Forum](#)

כיצד המאמרים קשורים למחקר שלנו ועוזרים לו:

- המאמרים מספקים רקע חשוב על ההשפעות הרחבות של מגפת הקורונה על תעשיית הספורט. הבנת ההשפעות הכלכליות והארגוניות מאפשרת לנו להבין את ההקשר הרחב שבו התקיימו טורנירי ATP בתקופת הקורונה.
- המאמרים מצביעים על כך שהמגפה גרמה לשיבושים רבים באירועי ספורט מרכזיים, דבר שהשפיע על דינמיקות המשחקים והביצועים של השחקנים. מידע זה חשוב למחקר שלנו, שכן הוא מסביר את הסיבות האפשריות לשינויים שמצאנו במדדי המרכזיות ובמבנה הקהילות בטורנירי ATP.

שאלת המחקר

כיצד השפיעה מגפת הקורונה על ביצועי השחקנים ותוצאות המשחקים בטורנירי טניס ATP ?
במחקר זה רצינו לבחון את השפעת מגיפת הקורונה על טורנירי טניס בעולם, וכך להבין את ההשפעה של מגיפה שהגיעה בהפתעה לעולם ולהבין כיצד ניתן להתמודד עם אירועים כאלה בעתיד. הדעות שלנו בנושא נחלקות לשני כיוונים שונים: האחד, אירועי הקורונה גרמו להפחתה במשחקים ולירידת רמת הביצועים של השחקנים. השני, השחקנים ניצלו את הזמן הפנוי להשקיע יותר באימונים, כלומר, הקורונה דווקא סייעה להם להשתפר ולהעלות את רמת הביצועים שלהם.

הכנת הנתונים

את כל הקוד של מהלך הפרויקט ניתן למצוא בקובץ ה-IPYNB המצורף להגשה.

הנתונים נאספו מאתר Kaggle וכוללים כ-60 אלף רשומות, כאשר כל רשומה מייצגת משחק טניס בין שני שחקנים בטורניר מסוים. הנתונים מכסים טורנירי טניס בין השנים 2000 ועד היום. אנחנו ביצענו סינון ובחרנו לעבוד עם השנים 2018-2021 בלבד כדי להתמקד בתקופות של מגיפה הקורונה והתקופה שלפניה.

חילקנו את המידע לשתי קבוצות – לפני מגפת הקורונה (2018-2019) ובמהלך תקופת הקורונה (2020-2021).

עמודות הדאטה כוללות תאריך, שחקנים המשתתפים במשחק, שם הטורניר, סוג המגרש, סיבוב בטורניר, ניקוד, דירוג כללי של השחקנים, מספר מערכות ששוחקו, תוצאת המשחק. הסרנו עמודות שאינן רלוונטיות לניתוח והוספנו עמודות שחישבנו מתוך הנתונים הקיימים לצורך הבדיקה במהלך הניתוח.

השלבים שביצענו לצורך הכנת הדאטה:

1. צמצום הדאטה לשנים 2018-2021.
2. ניקוי שורות עם ערכים ריקים.
3. תיקון שגיאות טכניות והגיוניות במערך הנתונים.
4. הסרת כפילויות או עמודות לא רלוונטיות לניתוח.
5. החלפת העמודות 1_Player ו-2_Player בעמודות Winner ו-Loser.
6. החלפת העמודות של נקודות ודירוג השחקנים, בהתאם לעמודות Winner ו-Loser.
7. הוספת עמודת Power_Rating המחושבת כמספר המערכות ששוחקו חלקי המספר המרבי האפשרי. הערכים יומרו לקטגוריות בין 0 ל-1.

בניית הגרפים

בנינו שני גרפים המייצגים שתי תקופות: תקופה הקורונה 2020-2021 והתקופה שלפניה 2018-2019.

השתמשנו בספריית NetworkX לבניית הגרפים. לכל צומת (שחקן) הוספנו את המאפיינים הרלוונטיים לניתוח על ידי חישוב ממוצעים מכל המשחקים שאותו שחקן שיחק בהם. מכיוון ששחקן מסוים עשוי לשחק במספר משחקים עם שחקנים שונים, חישבנו ממוצע כללי כדי שלכל צומת יהיו מאפיינים מתאימים ונכונים עבורה.

הגדרת צומת

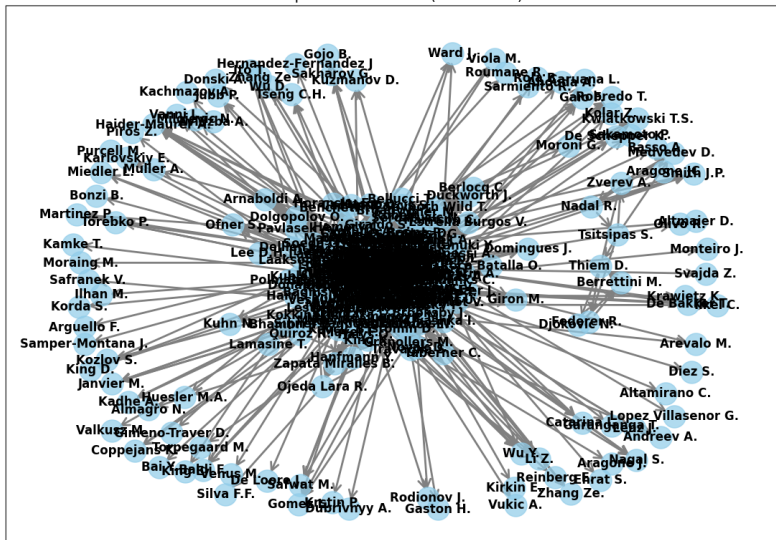
צומת מייצג שחקן יחיד בטורניר. כל צומת מכיל מידע על השחקן, כגון שם, ממוצע דירוגים, וממוצע נקודות במשחק.

הגדרת קשת

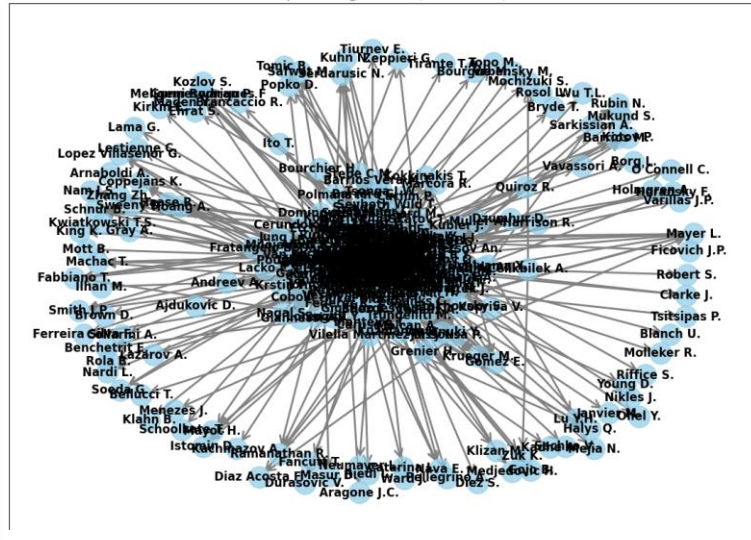
קשת מייצגת משחק שהתקיים בין שני שחקנים. הקשת מכוונת מהצומת של השחקן שניצח (Winner) אל הצומת של השחקן שהפסיד (Loser). בנוסף על כל קשת יהיה את המידע הבא:

Power_Rating: ככל שהניקוד גבוה כך השחקן ניצח בקלות יותר (0-1)
סוג מגרש: indoor/outdoor
חומר המגרש: קשיח/חמימרידשא

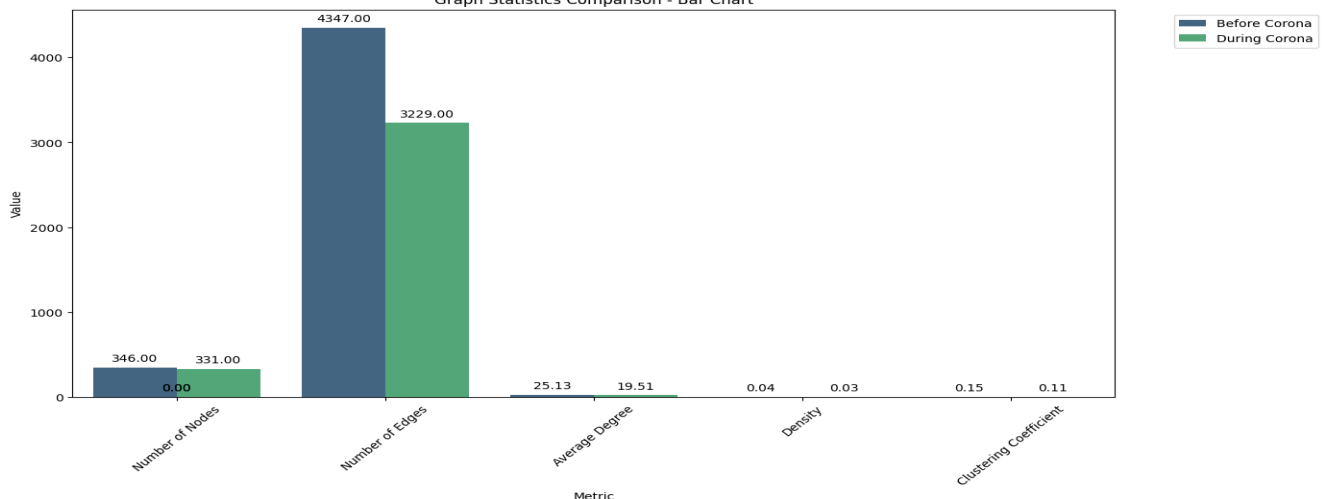
Graph Before Corona (2018-2019)



Graph During Corona (2020-2021)



Graph Statistics Comparison - Bar Chart



חישוב מדדים

לאחר בניית הגרף התחלנו בחישוב מדדים באמצעות ספריית NetworkX כדי לקבל מדדים עבור הרשת שלפני תקופת הקורונה והרשת במהלך תקופת הקורונה - כך שלבסוף נוכל להשוות בין מדדים אלה כדי להבין את ההשפעה של המגיפה והשינויים שקרו כתוצאה ממנה.

המדדים שחישבנו עבור כל רשת הם:

Degree Centrality

הסבר: מדד זה מחשב את מספר הקשרים הישירים שיש לכל צומת ברשת. כלומר, כמה משחקים שיחק כל שחקן. מדד זה חשוב למחקר שלנו משום שהוא מאפשר לזהות את השחקנים שהיו פעילים ביותר בטורנירים והאם הייתה ירידה במספר המשחקים שלהם במהלך המגפה.

תוצאות:

בתקופה שלפני הקורונה השחקנים הפעילים ביותר היו:

- עם Medvedev D. Degree Centrality של 0.301
- עם Zverev A. Degree Centrality של 0.295
- עם Tsitsipas S. Degree Centrality של 0.292

במהלך הקורונה השחקנים הפעילים ביותר היו:

- עם Rublev A. Degree Centrality של 0.266
- עם Auger-Aliassime F. Degree Centrality של 0.242
- עם Sinner J. Degree Centrality של 0.233

Closeness Centrality

הסבר: מדד זה מחשב את המרחק הממוצע של כל צומת מכל שאר הצמתים ברשת. כלומר, כמה קרוב כל שחקן לכל שאר השחקנים. מדד זה חשוב למחקר שלנו משום שהוא מאפשר לזהות את השחקנים שהיו בעלי גישה ישירה יותר לשחקנים רבים אחרים, דבר שמעיד על פעילות רחבה ומעורבות גדולה יותר בטורנירים.

תוצאות:

בתקופה שלפני הקורונה השחקנים הקרובים ביותר לאחרים היו:

- עם Struff J.L. Closeness Centrality של 0.342
- עם Sousa J. Closeness Centrality של 0.336
- עם Verdasco F. Closeness Centrality של 0.332

במהלך הקורונה השחקנים הקרובים ביותר לאחרים היו:

- עם Mannarino A. Closeness Centrality של 0.317
- עם Ramos-Vinolas A. Closeness Centrality של 0.316
- עם Basilashvili N. Closeness Centrality של 0.315

Betweenness Centrality

הסבר: מדד זה מחשב את מספר הפעמים שכל צומת נמצאת על המסלולים הקצרים ביותר בין זוגות של צמתים אחרים. כלומר, כמה פעמים כל שחקן משמש כמתווך בין שחקנים אחרים. מדד זה חשוב

למחקר שלנו משום שהוא מאפשר לזהות את השחקנים שהיו מרכזיים ביותר ברשת, כלומר שהיו חלק ממספר גדול של מסלולים קצרים בין שחקנים אחרים, דבר שמעיד על חשיבותם ברשת.

תוצאות:

לפני הקורונה:

- Simon G. עם Betweenness Centrality של 0.024
- Sousa J. עם Betweenness Centrality של 0.023
- Struff J.L. עם Betweenness Centrality של 0.021

במהלך הקורונה:

- Delbonis F. עם Betweenness Centrality של 0.026
- Sonego L. עם Betweenness Centrality של 0.022
- Norrie C. עם Betweenness Centrality של 0.022

Out Degree Centrality

הסבר: מדד זה מחשב את מספר הקשרים היוצאים מכל צומת ברשת מכוונת. כלומר, כמה משחקים ניצח כל שחקן. מדד זה חשוב למחקר שלנו משום שהוא מאפשר לזהות את השחקנים שהיו המנצחים הגדולים ביותר בטורנירים והאם הייתה ירידה במספר הניצחונות שלהם במהלך המגפה.

תוצאות:

לפני הקורונה:

- Medvedev D. עם Out Degree Centrality של 0.214
- Tsitsipas S. עם Out Degree Centrality של 0.197
- Zverev A. עם Out Degree Centrality של 0.194

במהלך הקורונה:

- Rublev A. עם Out Degree Centrality של 0.193
- Djokovic N. עם Out Degree Centrality של 0.169
- Zverev A. עם Out Degree Centrality של 0.164

השוואת המדדים לפני הקורונה ובמהלך הקורונה מראה שינוי בדינמיקה ובפעילות השחקנים.

בכל חישובי המדדים שביצענו, ניתן לראות שיש שינוי משמעותי בין התקופות השונות. השחקנים המובילים במדדים בתקופה שלפני הקורונה כבר לא מובילים במדדים בתקופת הקורונה. דבר זה מעיד על שינוי משמעותי ביכולות של השחקנים, ירידה ביכולת, פחות משחקים ששוחקו ועוד במהלך תקופת הקורונה.

המסקנה שלנו מחישוב המדדים הללו היא שהקורונה השפיעה ושינתה את מעמד השחקנים במשחקי הטניס בטורנירים השונים, אך מפני שהשינוי דרסטי מאוד ולא ניתן להשוות שחקנים שנשארו במעמד גבוהה בחישוב המדדים אז החלטנו לחשב מדדים נוספים ולהשתמש באלגוריתמים לזיהוי קהילות שאולי יעזרו לנו למצוא הבדלים ודברים זהים בין התקופות כדי להסיק מסקנות טובות יותר.

Small World

הסבר: מדד זה מחשב את מאפייני "עולם קטן" של הרשת, כלומר את מקדם הצבירה (clustering coefficient) ואת אורך המסלול הממוצע ברשת. מדד זה חשוב למחקר שלנו משום שהוא מאפשר להבין האם הרשת הפכה ליותר או פחות מקושרת ומרכזית במהלך המגפה, דבר שמעיד על שינויים במבנה הרשת החברתית של השחקנים.

תוצאות:

לפני הקורונה, הרשת הייתה יותר מקושרת ומרכזית, עם מקדם צבירה גבוה ואורך מסלול ממוצע קצר. זה מעיד על כך ששחקנים היו מעורבים יותר בטורנירים והיו בקשר קרוב יותר אחד עם השני.

במהלך הקורונה, הרשת הפכה לפחות מקושרת ומרכזית, עם מקדם צבירה נמוך ואורך מסלול ממוצע ארוך יותר. זה מעיד על כך שהשחקנים היו פחות מעורבים בטורנירים והקשרים ביניהם היו פחות חזקים.

Scale-Free

הסבר: מדד זה מחשב את ההתאמה של הרשת לחוק החזקה (power law), כלומר האם הרשת מכילה מספר קטן של צמתים עם הרבה מאוד קשרים ורוב הצמתים עם מעט קשרים. מדד זה חשוב למחקר שלנו משום שהוא מאפשר להבין האם הרשת הפכה ליותר או פחות היררכית במהלך המגפה, דבר שמעיד על שינויים במבנה הרשת והדינמיקות שלה.

תוצאות:

לפני הקורונה המקדמים של חוק החזקה (power law) היו נמוכים יותר (0.059), מה שמעיד על כך שהרשת הייתה יותר היררכית. היו מספר קטן של שחקנים מרכזיים שהיו מאוד פעילים ברוב הטורנירים, ורוב השחקנים האחרים היו פחות פעילים.

במהלך תקופת הקורונה המקדמים של חוק החזקה (power law) היו גבוהים יותר (0.075), מה שמעיד על כך שהרשת הפכה לפחות היררכית. כלומר, הייתה חלוקה שווה יותר של קשרים בין השחקנים, והפעילות בטורנירים הייתה מחולקת בצורה שווה יותר בין השחקנים.

השינויים במדדי Small World ו-Scale-Free מצביעים על כך שמגיפת הקורונה השפיעה באופן משמעותי על מבנה הרשת החברתית של השחקנים בטורנירי הטניס. לפני הקורונה, הרשת הייתה יותר מקושרת ומרכזית, והיו מספר שחקנים מרכזיים שהשתתפו ברוב הטורנירים והיו מאוד פעילים. במהלך הקורונה, הרשת הפכה לפחות מקושרת ופחות היררכית, עם חלוקה שווה יותר של הפעילות בין השחקנים. המסקנה היא שהמגפה גרמה להפחתת המעורבות והפעילות בטורנירים, והשפיעה על המרכזיות והקשרים בין השחקנים, דבר שיכול להעיד על הפחתת התחרויות והירידה בכמות הטורנירים והמשחקים שהתקיימו במהלך תקופה זו.

זיהוי קהילות

לאחר שביצענו חישוב של מדדי מרכזיות והשוונו בין התוצאות של המדדים על תקופת הקורונה והתקופה שלפניה, החלטנו להשתמש באלגוריתמים שונים לזיהוי קהילות כדי לבחון את השינוי בקהילות בין שתי התקופות. השתמשנו בשני אלגוריתמים שונים כדי להבין את ההשפעות בשתי גישות שונות ולבחון את שתיהן.

Girvan-Newman

אלגוריתם Girvan-Newman מבוסס על הסרת קשתות עם מדד Betweenness Centrality גבוה, כלומר הקשתות שמשפיעות על קישוריות הרשת בצורה משמעותית. האלגוריתם מתחיל בהסרת הקשתות עם הערך הגבוה ביותר וממשיך עד שהרשת מתפרקת לקהילות. זה מאפשר זיהוי של מבנים מרכזיים ברשת, ונותן הבנה על האינטראקציות המרכזיות בין השחקנים.

Louvain

אלגוריתם Louvain מזהה קהילות על ידי מקסום מודולריות (Modularity), שהיא מדד לכמה טוב הרשת מחולקת לקהילות. האלגוריתם מתחיל בחלוקת הרשת לקהילות קטנות ומבצע מיזוגים חוזרים ונשנים של קהילות כדי למקסם את המודולריות הכוללת. האלגוריתם יעיל ומהיר במיוחד, ומתאים לרשתות גדולות ומורכבות.

למה השתמשנו בשני האלגוריתמים ומה התוצאות תורמות למחקר:

- **Girvan-Newman**: השימוש באלגוריתם זה מאפשר לנו לזהות את הקשרים המרכזיים והחשובים ביותר ברשת. זה מסייע להבין מי הם השחקנים המרכזיים ביותר ומה התפקיד של הקשרים המרכזיים ברשת. התוצאה מראה על שינוי במרכזיות הקשרים בתקופת הקורונה, מה שמצביע על שינוי במבנה הרשת.
- **Louvain**: השימוש באלגוריתם זה מאפשר לנו לזהות את הקהילות בצורה מהירה ויעילה, ולראות כיצד הרשת מחולקת לקהילות שונות בתקופות שונות. מודולריות גבוהה יותר מצביעה על חלוקה טובה יותר לקהילות. התוצאות מראות שבתקופת הקורונה הייתה חלוקה יותר ברורה לקהילות, מה שמעיד על כך שהשחקנים היו יותר ממוקדים בקהילות מסוימות.

התוצאות:

לאחר בחינה של שני האלגוריתמים הנ"ל החלטנו להשתמש באלגוריתם של Louvain כדי לבחון את התוצאות וממה השינוי במודולריות נובע. אלגוריתם Louvain מתאים לרשתות גדולות יותר ומסובכות מה שמתאים לרשת שלנו ולמבנה שלה.

אלגוריתם Louvain, המבוסס על אופטימיזציה של מודולריות, הראה עלייה במודולריות במהלך הקורונה מ 0.136 ל 0.17. תוצאה זו מצביעה על כך שהרשת הפכה ליותר מקושרת עם מבנה קהילתי יציב וברור יותר, מה שמעיד על חלוקה שווה יותר של קשרים בין השחקנים.

מסקנות:

העלייה במודולריות בתקופת הקורונה, כפי שהראה אלגוריתם Louvain, מעידה על כך שהרשת הפכה למובנית וברורה יותר עם קהילות יציבות וחזקות יותר.

השפעת הקורונה על השחקנים והמשחקים:

- פחות משחקים בין קהילות שונות: העלייה במודולריות מצביעה על ירידה במשחקים בין קהילות שונות ועל ריבוי משחקים בתוך קהילות.
- חיזוק הקשרים בתוך הקהילות: השחקנים שמרו על קשרים חזקים יותר בתוך הקהילות שלהם, כנראה בעקבות פחות טורנירים בינלאומיים.
- הפחתת הניידות: הגבלות נסיעות גרמו לשחקנים להשתתף פחות בטורנירים מחוץ לאזור שלהם, מה שחיזק את הקשרים המקומיים.
- יצירת קהילות חדשות: ייתכן שנוצרו קהילות חדשות שהתארגנו סביב טורנירים מקומיים ומסגרות משחק מצומצמות יותר.

סיכום ומסקנות

בפרויקט זה התמקדנו בניתוח השפעת מגפת הקורונה על ביצועי השחקנים ותוצאות המשחקים בטורנירי ATP. השתמשנו במדדי רשת ובאלגוריתמים לזיהוי קהילות כדי לבחון שינויים במדדי המרכזיות והקהילות של השחקנים לפני ובמהלך המגפה. הממצאים מצביעים על ירידה בפעילות ובקישוריות של שחקנים מובילים, ומספקים תובנות מעניינות על השפעת המגפה על מבנה הרשת החברתית של השחקנים.

מסקנות עיקריות

1. ירידה בפעילות של שחקנים מובילים:
 - נצפתה ירידה במספר המשחקים ובניצחונות של שחקנים מרכזיים. סביר להניח כי מגבלות התנועה והביטול של טורנירים רבים השפיעו על זמינות המשחקים ועל התחרויות.
2. פחות קשרים ישירים:
 - הירידה במספר המשחקים הביאה לירידה בקשרים הישירים בין השחקנים. מגבלות הנסיעות והביטולים הרבים תרמו להפחתה בקישוריות בין שחקנים מטורנירים שונים.
3. שינוי במבנה הקהילות:
 - המבנה הקהילתי של השחקנים השתנה במהלך המגפה, עם היווצרות קהילות ברורות וממוקדות יותר. השחקנים התמקדו בקהילות קטנות ומבודדות יותר, מה שהוביל לפיזור אחיד יותר של הקשרים בין השחקנים.
4. פחות היררכיה ברשת:
 - נצפתה ירידה בהיררכיה ברשת החברתית, עם חלוקה שווה יותר של קשרים בין השחקנים. נראה כי המגפה הובילה לחלוקה יותר מאוזנת של הקשרים ברשת.

דיון

מגפת הקורונה השפיעה באופן משמעותי על מבנה הרשת החברתית של השחקנים בטורנירי ATP. היא גרמה לירידה בפעילות ובקישוריות של שחקנים מרכזיים, והובילה לשינויים במבנה הקהילות עם חלוקה ברורה ושווה יותר של קשרים בין השחקנים. הממצאים ממחישים את ההשפעות של אירועים גלובליים על דינמיקות המשחקים והמבנה החברתי בטורנירים.

מסקנות נוספות

נראה כי מגיפת הקורונה השפיעה על מדדים רבים ושינתה את הקישוריות בין השחקנים. מהממצאים ניתן להבין שלאירועים מהסוג הזה ישנה השפעה רבה על עולם הספורט ובפרט על טניס, מפני שבענף זה יש קשרים רבים בין שחקנים.

המלצות להמשך

לאור השינויים המשמעותיים שהביאה המגפה, אנחנו ממליצים על הכנה עתידית למניעת שינויים דרסטיים במקרה של אירועים דומים בעתיד. יש להתכונן ולדעת איך להכין את השחקנים במצבים כאלו, וכיצד להמשיך לקיים משחקים במתכונת שונה במקום לבטלם לחלוטין. שמירה על תפקוד רגיל עד כמה שניתן היא קריטית לשמירה על מבנה הרשת החברתית והקשרים בין השחקנים.