# Contents

1 Mathematical Models of Radicalization and Terrorism ............... 1  
Yaoli Chuang and Maria R. D’Orsogna  
1.1 Introduction ....................................................... 2  
1.2 Radicalization in opinion dynamics models ....................... 8  
1.3 Radicalization in aging models ................................. 13  
1.4 Lattice and network models .................................. 16  
1.5 Game theoretic models of terrorism ............................ 28  
1.6 Terrorist events as self-exciting processes .................... 35  
1.7 Summary and conclusions .................................... 39  
1.8 Acknowledgments .............................................. 40  
References ............................................................. 40
Chapter 1
Mathematical Models of Radicalization and Terrorism

Yaoli Chuang and Maria R. D’Orsogna

Abstract The rapid spread of radical ideologies has led to a world-wide succession of terrorist attacks in recent years. Understanding how extremist tendencies germinate, develop, and drive individuals to action is important from a cultural standpoint, but also to help formulate response and prevention strategies. Demographic studies, interviews with radicalized subjects, analysis of terrorist databases, reveal that the path to radicalization occurs along progressive steps, where age, social context and peer-to-peer exchanges play major roles. To execute terrorist attacks, radicals must efficiently communicate with one another while maintaining secrecy; they are also subject to pressure from counter-terrorism agencies, public opinion and the need for material resources. Similarly, government entities must gauge which intervention methods are most effective. While a complete understanding of the processes that lead to extremism and violence, and of which deterrents are optimal, is still lacking, mathematical modelers have contributed to the discourse by using tools from statistical mechanics and applied mathematics to describe existing and novel paradigms, and to propose novel counter-terrorism strategies. We review some of their approaches in this work, including compartment models for populations of increasingly extreme views, continuous time models for age-structured radical populations, radicalization as social contagion processes on lattices and social networks, agent based models, game theoretic formulations. We highlight the useful insights offered by analyzing radicalization and terrorism through quantitative frameworks. Finally, we discuss the role of institutional intervention and the stages at which de-radicalization strategies might be most effective.

Yaoli Chuang
Department of Mathematics, California State University at Northridge, 18111 Nordhoff Street, Los Angeles, CA 91330 e-mail: yaoli.chuang@csun.edu

Maria R. D’Orsogna
Department of Mathematics, California State University at Northridge, 18111 Nordhoff Street, Los Angeles, CA 91330e-mail: dorsogna@csun.edu
Fig. 1.1 Terrorist attacks between 1970-2015, from the Global Terrorist Database (GTD) as compiled by the National Consortium for Study of Terrorism and Responses to Terrorism (START). The GTD details more than 180,000 events including bombings, assassinations, and kidnappings. Data is entered upon reviewing news articles, court reports and other sources. Among the most notable groups the FARC in Colombia, Shining Path in Peru, and the FMLN in El Salvador; the Red Brigades in Italy, the Red Army Faction in Germany, the PIRA in Northern Ireland and the ETA in the Basque country; Boko Haram in Nigeria and Al-Shabaab in Somalia; ISIS, Al-Qaeda and the Taliban in the Middle East; the LTTE and the NPA in South and Southeast Asia, as well as several groups in Jammu and Kashmir.

1.1 Introduction

The terrorist attacks of September 11th 2001 destroyed lives and buildings, inflicting pain on many innocent families. Almost twenty years later, they have also profoundly changed our lives, from the mundane aspects of boarding an airplane to the definition of war, peace, tolerance. In their aftermath it is no longer shocking to view a line of code, a truck, a pair of shoes, or an aircraft as part of a sinister threat. At the same time, in the name of security and safety, we have witnessed the erosion of some of our civil liberties, and we have come to accept that the privacy of our email exchanges, computer log-ins, financial transactions, and travel records may no longer be sacrosanct. One of the most troubling aspect of the September 11th events is that they instilled fear in the population, contributing to greater political polarization and divisiveness that continue to this day, periodically fueled by segments of the media and political punditry. Of course, this is the goal of terrorism, which in a nutshell is a political and psychological strategy to weaken the “enemy” from within.

While on an unprecedented scale, the events of September 11th are not isolated: terrorism, radicalization and the glorification of violence are worldwide phenomena that aim to disrupt civilian life and threaten security. In Figure 1.1 we show global terrorist attacks between 1970 and 2015 as compiled by the National Consortium for Study of Terrorism and Responses to Terrorism [1]. As can be seen, many corners of the world are hotspots for terrorist activity, inspired and driven by a wide spectrum
of motivations or ideologies that can be religious, political, nationalist or subversive. Extremists exploit traditional and, in an increasingly globalized world, online tools to disseminate propaganda, recruit new adherents, plot attacks [2]. They do so often unhindered by geographical boundaries. The question of how to guarantee an open society, freedom of expression, belief, movement while at the same time preventing violence, is an unresolved one, dating all the way back to classical antiquity [3].

A terrorist event does not occur out of the blue. Individuals who commit these acts have typically spent months or years “radicalizing”: reading fanatical sources, discussing with, or emulating, others with more extreme views, becoming proactive and later orchestrating the attack with like-minded peers. Numerous sociological, psychological and economic studies have attempted to understand the overall process of radicalization, through demographic studies, by interviewing subjects, by analyzing socio-economic backgrounds and modus operandi of extremists [4, 5, 6, 7, 8, 9, 10, 11]. One example is the Profiles of Individual Radicalization in the United States (PIRUS), a database of domestic Islamist, Far Left and Far Right radicalized individuals [12]. Distinctions have been made to differentiate when ideological radicalization stays part of a belief system, however contorted, and when it leads to bloodshed [13, 14, 15]. What has emerged is that while some general trends can be outlined, there is no one single, key element that can explain why a person, or a group of people, choose to reject the status quo and embrace indiscriminate violence [16].

The making of a radical typically unfolds in a multi-step fashion and is strongly influenced by one’s surroundings, past experiences and future prospects [17, 18, 19, 20, 21, 22]. Depending on the particular socio-geographical context, four or five main steps have been identified: pre-radicalization (at risk), self-identification (susceptible), indoctrination (moderate radical), commitment and “jihadization” (full radicalization) [20, 23, 24]. More nuanced classifications, expanding these classes to eight states, have also been proposed [25], while the IVEE theory focuses on three main steps: individual vulnerability (IV), exposure (E), emergence (E) [26, 27]. The time-scale for individuals to advance through these hierarchies varies, but case studies conducted on sixty-eight American homegrown Al-Qaeda inspired radicals reveal that pre-radicalization takes four to five years, while progression through the following stages is faster and occurs over a timeframe of less than three years [28, 29, 30, 31]. Other studies conducted on ten religious radical groups by New York Police Department detectives, analysts and intelligence officials show that the time spent in the at risk or in the susceptible class is approximately one and a half and three years, respectively; the time in the more radical stages is about two years each [20]. Overall, radicalization times are becoming shorter due to more rapid information sharing, especially online.

The seed of pre-radicalization is often rooted in various unaddressed grievances and personal frustrations, such as lack of employment and opportunity, racial and religious discrimination and social exclusion [32, 33, 34]. Individual malaise can also stem from real or perceived socio-economic injustices, against policies that are seen as too progressive, or conversely, because of the desire for sweeping societal change. Personal dissatisfaction leads to self-identification, where marginalized
or self-searching individuals gradually begin constructing new identities and routines, shifting away from old ones. Like-minded people are actively sought and new friendships formed [35]. The process is furthered by indoctrination when bonds of solidarity within these newly formed groups strengthen. Mutual encouragement and the absence of a counter-dialectic allow for extremist views to self-reinforce and become more engrained [36]. Once embraced, disavowing the new ideals becomes difficult; justification and praise of violence follow [37]. A further activism phase ensues, when radicals commit to militantly spread their convictions to others until external events such as political or judiciary decisions, or simple serendipity, crystallize the willingness to take violent action [38, 39]. A schematic of this process is shown in Figure 1.2.

Radicalization typically occurs through networks of peers and may be facilitated through technology in the form of web-based recruitment material or chat rooms [40, 41]; ISIS is a well known internet savvy group, that uses social media to recruit western foreign fighters in Syria and Iraq [42, 43]. A recent RAND study shows that the virtual world offers extremists the capability to communicate, collaborate and convince without physical contact and enables connection with like-minded individuals from across the world at all times [44]. It also may act as a modern day echo-chamber to confirm existing beliefs.

Extremist tendencies can arise at any age and factors that are traditionally associated with desistance from criminal or deviant behavior, such as marriage, career and education, are not necessarily strong deterrents against ideological extremism [4, 45]. Disenfranchised adolescents and young adults, however, are particularly vulnerable to indoctrination and radicalization, especially when their formative years
Fig. 1.3 Estimated age of interviewed subjects at the first indication of radicalization. All were radicalized within the United States, inspired by Al-Qaeda or ISIS, and committed terrorist acts over a period of 16 years after September 11th 2001. Group B (129 individuals) is a subset of Group A (289 individuals) for which more detailed data was available. Breaks indicate that no subjects of that age cohort were included in the study. Although individuals can be radicalized at any age, the process is more common during adolescence and early adulthood. Taken from Ref. [31].

are spent without purpose, education, outlook or positive role models [38]. This is often the case for marginalized youth, religious zealots, or right-wing extremists [46, 47] for which radicalism may provide a sense of purpose. Figure 1.3 shows that the majority of 289 homegrown terrorists in the United States, inspired by Al-Qaeda and ISIS, initiated their radicalization process before age 25 [31]. Although in their infancy and mostly experimental, de-radicalization programs are being developed and implemented worldwide with great expectations. They include education, exposure to literature, sports and arts, psychological counseling, job training and informal one-on-one conversations [48]. The goal is to prevent radicalization, disengage known violent extremists, and reintegration former convicted extremists in the community [49, 50, 51, 52, 53, 54, 55]. Intervention through social media has also been advocated: it is hoped that the interactiveness afforded by the internet can be used to help change the perceptions of budding radicals [56].

It may be natural to ask whether parallels exist between radicalization and the process by which a propensity for crime develops. Indeed, some studies have found similarities between becoming an extremist and joining urban street gangs [57, 58]. Several observations must however be made in this respect. Terrorists or groups of terrorists take action much less frequently than urban criminals; radicals are driven by deeply held ideologies and not by opportunistic, short-term rewards; selecting the appropriate times and locations, and preparing for a terrorist attack, are much more elaborate processes that committing street-level crimes; and finally terrorists
typically seek worldwide attention in the aftermath of their acts, whereas burglars and other criminals prefer anonymity. Of course, the outcomes of successful terrorist and criminal acts are also much different in terms of damages inflicted to those directly affected and to society as a whole [59, 60, 61]. Hence, although some comparisons can be made, radicalizing and becoming an urban criminal, should be considered distinct phenomena.

Once the radicalization process is complete, a group of like-minded extremists may decide to turn their fanatical ideology into violent action. This requires willingness, but also training, preparation, curating the logistics. Eventually, organizational structures emerge, either organically or by design. The way individuals connect depends on specific needs and choices such as targets, ideology, timing, resource availability, governmental efficiency, public opinion [15, 62]. One of the most important determinants of terrorist connectivity is the need to balance security and efficiency: on one hand, extremists must communicate and plan, on the other, they must protect themselves against infiltration, threats and counter-terrorism intervention. Governmental agencies must also decide the best course of action to disrupt these “dark” networks, be it via direct repression, manipulation, or persuasion, and by considering several environmental variables, such as costs, public perception, odds of success, information at hand. The final goal may be to splinter and dissolve the terrorist network, to weaken its financial and societal support structure, or to moderate its use of violence [63]. Of course, counter-terror tactics may prove to be counterproductive if they cause backlash and over-reaction in the population, so that policies must be calibrated to be firm but not excessive or discriminatory.

At present, two main counter-terrorism schools of thought exist. The Anglo-Saxon approach focuses on the rule of law so that only if, and after, violent acts are committed, security agencies will intervene. There is no curtailing of the beliefs these acts may stem from, in the name of freedom of belief and religion. The continental European approach instead attempts to prevent violence by confronting extremist ideas, especially the preaching of intolerance, before any attacks materialize [64, 65, 66]. Both have merits and drawbacks, and are deeply rooted in the history and national identity of the respective countries where they are applied. Finally, recent studies in political terrorism have identified “waves” of terrorism as 40-year generational cycles over which radical tendencies tend to ebb and flow [67]. Periods of heightened activity are characterized by the emergence of terror groups that share ideology, tactics, and vision for the future, that later subside. The wave theory of modern terrorism was crafted by David C. Rapoport [68] who reviewed terrorist events from the late 1800-s to the present day; it shares parallels with historian Arthur Schlesinger’s theory of 40-year generational cycles in politics [69].

The many processes and theories described above define complex, layered systems that involve the individual, the external socio-economic environment in which he or she is immediately placed, the connectivity of said individual through peer relationships and the larger cyber-world. All these elements, and the way they relate to each other, change in time due to prescribed dynamics, actions taken or not taken, and external perturbations. Certain trends are reinforced, others hampered,
occasionally leading to irreversible or unexpected outcomes depending on the particulars of the specific individual, community or system under investigation.

Unraveling the complexities of any social or behavioral phenomenon is not an easy task, especially when its is spread across generations, socio-economical conditions and geography. When seeking for general trends, often contradictory findings emerge due to the many factors affecting given observations that make each moment in space and history unique. Despite the many limitations imposed by the clandestine nature of extremist operatives, recent years have seen great advances in the collection, data-mining and statistical analysis of terrorist and radical data, such as profile lists of known extremists, databases listing terrorist attacks worldwide, the dissecting of twitter accounts, text analysis, the cataloguing of relevant hashtags and applying social network analysis [70, 71, 72, 73, 74, 75, 76, 77, 78, 79]. Delving into the data allows us to find features and patterns, to uncover intriguing irregularities, and spurs our curiosity. However, seldom does it lead to the identification of the underlying mechanisms causing the observed trends. Mathematical models offer an alternate path. Since they are based on observations that can be described by modifiable dynamical processes, initial conditions and parametric variables, mathematical models allow us to analyze the effects of individual inputs, the interplay between them, the effects of possible feedbacks and behaviors at different time scales. By using these inputs as building blocks, one can gradually build a nuanced understanding of the complex system at hand, identify emergent behavioral or spatio-temporal patterns, test intervention methods, identify the effects of constraints. Of course, measurability and quantifiability remain an issue and one must never forget that although useful in offering insight, no model can capture the full complexity of the real world. A relatively new approach is to conduct laboratory experiments, whereby participants are placed in virtual scenarios and must interactively select actions to take. They are then polled on their intents and motivations [63, 80]. Perhaps, the future lies in combining data-mining and other statistical analysis of compiled records with model-building, using tools from physical and mathematical sciences.

In this work, we review recent models of radicalization where statistical physics, complexity science, game and network theory are used to understand the multiple aspects involved and to identify possible prevention and amelioration strategies. A considerable body of work on mathematical models used for societal dynamics is already present in the literature [81, 82, 83], as applied to opinion, rumor and voter dynamics [84, 85, 86, 87, 88, 89, 90, 91], linguistics [92], migration and crowd behavior [93, 94, 96], crime and conflict [97, 98, 99, 100, 101, 102, 103, 104], social unrest [105] recidivism and rehabilitation [106], culture dynamics [107, 108, 109, 110], human cooperation [111, 112], warfare dynamics [113, 114]. Here, we specifically focus on the radicalization of a cohort of interacting individuals and on the dynamics of terrorist activity and intervention methods.
1.2 Radicalization in opinion dynamics models

We begin with radicalization models framed within the context of opinion formation, a most natural starting point since radicals can be viewed as individuals with extreme “opinions”. Here, the belief of an individual – and of society at large – change through peer-to-peer contact, the influence of media, or current events. Mean-field models assume that individuals with a given belief behave uniformly and define a homogeneous population; each of these populations is assumed to interact with others influencing their views, in a reversible or irreversible manner, so that as time evolves the size of a given population may increase or decrease. Mathematically, the dynamics is described by ODEs or, in some cases, by PDEs to include spatial dependence. These are simplistic models, that do not take into account the much larger complexities that lure single individuals towards fanaticism, yet they allow for the identification of parameters or mechanisms that drive observed trends.

The work of Carlos Castillo-Chavez [115] is one of the first to introduce population dynamics models to the study of transmission of fanatical behaviors, building on epidemiology contact processes [116]. Concepts such as “reproduction numbers” \( R_0 \) used to predict when a disease becomes endemic (\( R_0 > 1 \)) and when it does not (\( R_0 < 1 \)) are adapted to the spread of radical ideologies. In particular, a so called “fanatic hierarchy” is introduced with the total population \( T(t) \), divided into a subgroup \( G(t) \) that has no propensity for radicalization, and three subgroups of individuals at various stages of their commitment to the extreme ideology. The latter are \( S(t) \), a group of susceptibles who are not yet radicalized but who are vulnerable and open to the radical ideology at hand; the population \( E(t) \), individuals who have recently turned into fanatics; and finally \( F(t) \) individuals who have completely espoused radical views. The various subpopulations are related by \( T(t) = G + S + E + F \), and the model is written as

\[
\begin{align*}
\frac{dG}{dt} &= \Lambda - \beta_1 G \frac{C}{T} + \gamma_S S + \gamma_E E + \gamma_F F - \mu G, \\
\frac{dS}{dt} &= \beta_1 G \frac{C}{T} - \beta_2 S \frac{E}{C} - \gamma_S S - \mu S, \\
\frac{dE}{dt} &= \beta_2 S \frac{E}{C} - \beta_3 E \frac{F}{C} - \gamma_E E - \mu E, \\
\frac{dF}{dt} &= \beta_3 E \frac{F}{C} - \gamma_F F - \mu F, 
\end{align*}
\]  

(1.1)  
(1.2)  
(1.3)  
(1.4)

where \( C(t) = T(t) - G(t) \) represents the collection of individuals who are susceptible, partially or fully radicalized. The system includes a birth term \( \Lambda \) into the non-susceptible population \( G \) and a death rate \( \mu \) for all subpopulations, so that \( dT/dt = \Lambda - \mu T \), yielding \( T(t \to \infty) = \Lambda / \mu \). All other terms in Eqs.1.1–1.4 are associated to specific transitions between the various subpopulations. For example, individuals of the non-susceptible population \( G \) may become receptive to radical ideologies if exposed to them; the push of the radical message is modeled via the
Fig. 1.4 Schematics of the radicalization process according to representative compartment models [115, 117]. In panel (a) individuals may progress from a general, non-radical state $G$ towards a hierarchy of susceptibles $S$, recent adherents $E$ and full fanatics $F$. Birth and death are also included. In panel (b) two radical groups $\{E_1, F_1\}$ and $\{E_2, F_2\}$ can originate from the susceptible cohort $S$. The two ideologies interact with adherents moving between recent convert groups $E_1, E_2$.

The parameter $q < 1$ indicates how effective groups are at cross-recruitment. The transition $E_2 \rightarrow E_1$ is modulated by the the full radical cohort $F_1$ and by a percentage of recent converts $qE_1$; the $S \rightarrow E_1$ transition is modulated by $E_1 + F_1$ and by some members of the opposite group $qE_2$ who proselytize in favor of their counterparts. Similar arguments hold for $E_1 \rightarrow E_2$ and $S \rightarrow E_2$. In both panels, $T$ is the total population and $C = T - G$.

population ratio $C/T < 1$ modulated by the rate $\beta_1$ so that the net flow $G \rightarrow S$ is expressed by $\beta_1 G C / T$. Similarly, the flow from the susceptible state into the first stages of fanaticism $S \rightarrow E$ depends on the indoctrination and on the example set by the more radical individuals and is given by $(E + F)/C$ modulated by the rate $\beta_2$. The transition from the initial radical state to the fully committed one $E \rightarrow F$ is instead driven solely by the radical cohort via $F/C$ and modulated by the rate $\beta_3$. The model is summarized in Fig. 1.4(a).

Eqs. 1.1–1.4 define a hierarchy in the sense that transitions through the $G \rightarrow S \rightarrow E \rightarrow F$ stages are propelled by more radical populations influencing the lesser ones. The three rates $\gamma_i$, for $i = 1, 2, 3$, represent de-radicalization from the susceptible, recent convert, and fully fanatical subgroups, respectively. Note that there are no
reverse intermediate transitions: radical individuals $F$ for example may return to the non-susceptible population $G$ but do not transition back to their first stage of radicalization $E$.

Despite its simplicity, the model in Eqs.1.1–1.4 offers useful insight on societal outcomes, in terms of attractor points and thresholds. Of course the most important question to ask is under which conditions does a finite fanatic population $F$ arise. Analysis of Eqs.1.1–1.4 shows that no level of radicalization will be sustained ($S = E = F = 0$) for $\gamma_1 > \beta_1$, implying that one way to avoid radical discourse in a society is to hamper the onset of the radical process at the early $G \rightarrow S$ stage, when individuals become susceptible to extremism. Depending on other parameter combinations, attractor points with no fanatic populations ($S^* \neq 0, E = F = 0$) or ($S^* \neq 0, E^* \neq 0, F = 0$), and with fanatic populations ($S^* \neq 0, E^* \neq 0, F^* \neq 0$) can be identified. Equally important are the initial conditions and time scales involved: for example, even if the steady state is predicted to yield $F = 0$, an initially small cohort of extremists can successfully invade the population and lead to a large fanatic population over intermediate times, before it begins to decay.

Numerous studies have built upon Eqs.1.1–1.4. The possibility of two competing radical groups emanating from the same general, non-susceptible population $G$ has been modeled by including two recruitment rates $q\beta_1$, and $(1-q)\beta_1$ feeding into two distinct susceptible groups [117]. Each of these two cores follows the same radicalization process illustrated in Eqs.1.1–1.4 without interacting. More interestingly, the two subgroups may also experience cross-contact across the $S \rightarrow E \rightarrow F$ hierarchy as illustrated in a different variant of the model shown in Fig. 1.4(b). Interactions between the two radical branches represent recruitment competition or inadequate retention efforts so that fanatics may draw new adherents from converts to the opposite ideology. Trade-offs between recruitment and retention are addressed, and parameter regimes whereby one of the two ideologies will persist and the other become extinct, are identified. In particular, it is shown that the two groups cannot both coexist at steady state without cross interactions: competitive recruitment is necessary for the emergence of two finite fanatic populations, $F_1, F_2$. Successive studies include simplified models that focus on the recruitment process outlined in [115, 117], that include de-radicalization treatments of extremists [118, 119], or that parameterize government intervention [120].

Compartment models akin to [115, 117] have also been applied to actual extremist movements, for example to study the influence of separatist groups in the Basque country of Spain [121, 122]. Subpopulations were created and transition parameters were estimated using demographic and electoral data; immigration and emigration terms were included as well. The ideological evolution of society was thus analyzed over a 35-year period, including projections into the future. Eqs.1.1–1.4 were also adapted to the study of radicalization of extreme right wing movements in Germany [123] where Likert scale surveys repeated biannually from 2002 onwards [124, 125] were used for calibration. In a separate study, data on ten violent extremist online forums was collected over four years to develop a visitor-engagement model [126]. It was found that users undergo several layers of involvement until they become recruiters or real-life terrorist operatives. The authors discuss strategies to control the
proliferation of online groups, and find that aiming for the dissolution of a large number of extremist cyber-communities is not very effective as enforcement costs increase but gains are limited. Targeted, occasional intervention towards large groups is a preferable policy, as it will induce other, minor, forums to self-regulate.

Alternate compartment models define subpopulations along cultural lines, so that responses to given societal issues may fuel radicalization in more sensitive subgroups [127]. Here, the core subgroup represents the mainstream; two sensitive subgroups include first to third generation immigrants whose ways of life may be partially incompatible with the core. Once conflict arises, core agents are assumed to be inflexible; the sensitive subgroups may choose to adjust to mainstream beliefs and attitudes, or to oppose them. Three categories thus arise: core-inflexible $\sigma_I$; sensitive-peaceful, $\sigma_P(t)$, and sensitive-opponent $\sigma_O(t)$. The latter subgroup is assumed to be the source of radical activity. Sensitive individuals can change their status depending on interactions among themselves as well as with the core, according to the following dynamics

\[
\frac{d\sigma_P}{dt} = \alpha\sigma_I\sigma_O(t) - \beta\sigma_O\sigma_P, \tag{1.5}
\]

\[
\frac{d\sigma_O(t)}{dt} = -\frac{d\sigma_P}{dt}, \tag{1.6}
\]

where the $\sigma_i$ values, $i = \{I, P, O\}$ represent subpopulation fractions so that $\sigma_I + \sigma_P(t) + \sigma_O(t) = 1$. Since the core is inflexible, the value of $\sigma_I$ is fixed. In Eqs. 1.5-1.6 opponents recruit peaceful individuals but this process is hindered by the core that tends to steer opponents towards the peaceful state. The authors investigate the conditions under which a small initial minority of opponents persists, leading to long term radicalization. Of the two steady states that arise, only one has a finite opponent population. In particular, for a large enough core $\sigma_I > \beta/(\alpha + \beta)$, opponents $\sigma_O(\infty) \to 0$, precluding radicalization. Although Eqs. 1.5-1.6 are limited to exchanges between segments of society and do not include institutional roles, the results highlight that one of the ways to thwart radicalization is through core citizens ($\sigma_I$) actively engaging with sensitive agents ($\sigma_O, \sigma_P$). The inflexible core $\sigma_I$ places this work among the many opinion dynamics models where a “zealot” fraction of the population holds a fixed view that can never be changed, and where final outcomes depend on its magnitude [88, 128].

Other compartment models have included possible catalysts for radicalization such as unpopular political decisions or foreign interventions [129]. These may generate hatred that spreads within a population, leading to outbreaks of violence. For example, in [129], five possible states describe societal attitudes towards a given issue or decision; the population can be ignorant but sensitive to the topic at hand, and once aware, immune, upset, or violent. All five groups influence each other and contagion unfolds from an initial condition of a small upset cohort interacting with those who are susceptible or immune to the news. After being upset and violent for some time, individuals may accept the status quo and become relaxed. Particular attention is given to the growth of the violent cohort whose ranks increase as the upset
population interacts with the rest. Finally, compartment models have also been used to study the internal dynamics of terrorist cells, with interacting groups of leaders and foot-soldiers; counter-terror measures targeting both groups are also included in the form of distinct death rates [130].

Following a slightly different approach, the emergence of two antagonistic, radicalized groups was modeled in [131]. Society is here assumed to host two competing religious, ethnic or political groups, each of them harboring a moderate and a radical faction that interact with each other. Included in the model is the possibility of one group violently attacking the other. Intra-group transitions within a sect (from moderate to radical, and vice-versa) are assumed to occur either spontaneously or through indoctrination, but also depend on the actions and characteristics taken by the other sect. This is done by including a sensitivity parameter that indicates how strongly individuals will react to external attacks or propaganda. The two A and B sects are divided into time dependent $r_A(t), r_B(t)$ radical and $n_A(t), n_B(t)$ moderate factions. Since the total population per sect is assumed to be constant, and $n_A(t) + r_A(t) = N_A$ and $n_B(t) + r_B(t) = N_B$, the time evolution of each sect is completely determined from the dynamics of the radical component. For sect A thus

$$\frac{dr_A}{dt} = n_A \left[ \lambda_A s_A + p_A s_A \frac{r_A}{N_A} - \mu_A f(s_A) r_A \right], \quad (1.7)$$

while sect B follows the same dynamics with $A \rightarrow B$ and vice-versa. The first term on the right-hand side of Eq. 1.7 represents spontaneous radicalization described by the intrinsic rate $\lambda_A$, modulated by the sensitivity parameter $s_A$. The second term represents radicalization in response to indoctrination from the $r_A/N_A$ fraction of active radicals at rate $p_A$, and similarly modulated by $s_A$. De-radicalization is assumed to occur at an intrinsic rate $\mu_A$ modulated by the sensitivity-dependent function $f(s_A)$. The latter is decreases with $s_A$ so that a highly sensitive population is more unlikely to de-radicalize. Finally, sect A is assumed to attack sect B at rate $k_A r_A$, where $k_A < \omega_B$, and $\omega_B$ is a maximal threshold. Coupling between the A and B sects occurs through the sensitivity parameter $s_A$. If this quantity is selected to be a numerical value, the two factions are effectively uncoupled; otherwise $s_A = k_B r_B / \omega_B N_A$ is assumed. The above relationship for $s_A$ implies that the sensitivity depends on the ratio between the number of attacks per unit time the $r_B$ radicals actually impart on sect A, via the term $k_B r_B$, and the hypothetical maximal attack rate sect A could inflict on sect B in retaliation, $\omega_B N_A$. In this context, sensitivity is a measure of the relative aggression capabilities of the two groups. The dynamics that unfolds from Eq. 1.7 is rich: steady states, stability and bifurcations are determined under several scenarios of sect size and interactiveness. A game theoretic framework [132, 133, 134, 135] is further added, assuming that radical factions may tune “strategic” parameters, such as intensity of propaganda $\lambda_{A,B}$ or frequency of attacks $k_{A,B}$ to optimize given utility functions aimed at increasing rank numbers while decreasing rival attacks. Resources are assumed to be limited so that choices must be made between proselytizing and attacking; a sect can also readjust its choices in response to the actions taken by the other. The main finding of this work is that unless very high rates of
violence are employed, the existence of relatively small groups of radicals cannot be sustained: eventually, they become less extreme and only the moderate faction will persist. The perspective that emerges is that to survive, radical groups must employ greater violence in their early days, when they are still numerically small, and may transition towards less violent methods, such as indoctrination of moderates, later on, as they mature.

Finally, system dynamics approaches to understand how transnational terrorists groups recruit new members, train and sustain them, and execute terrorist attacks have also been introduced, in particular to study the activities of Al-Qaeda and responses from the United States [136]. Non-linear behaviors are analyzed via so-called stocks and flows, feedback loops and time delays, allowing for several scenarios to be explored through graphical user interfaces [137].

### 1.3 Radicalization in aging models

The mathematical models described above provide many useful insights, however none include age-sensitive responses to propaganda, emulation of peers and societal pressure. As described in the introduction, radicalization is highly age-dependent, and age stratified models may help better understand the establishment and evolution of radical groups over time. A compartment model of increasingly fanatic stages coupled to age-differentiated interactions was introduced in [139]. For simplicity, only three cohorts are considered in this work: non-radical \( i = 0 \), activist \( i = 1 \) and radicals \( i = 2 \) described by the densities \( \rho_i(t, a) \) of age \( a \) at time \( t \) for \( i = \{0, 1, 2\} \).

Transitions between successive pools are mediated by the activation rate \( A(a; \rho_1) \) (from \( i = 0 \) to \( i = 1 \)) and the radicalization rate \( R(a; \rho_2) \) (from \( i = 1 \) to \( i = 2 \)). Both rates depend on the age of the source population as well as on the structure of the sink population. A de-activation rate \( C_D \) (from \( i = 1 \) to \( i = 0 \)) and a pacifying rate \( C_P \) (from \( i = 2 \) to \( i = 1 \)) are also included and assumed to be age-independent for simplicity. The resulting age-structured model is of the McKendrick-von Foerster type [140, 141, 142, 143, 144, 145] and is written as

\[
\begin{align*}
\frac{\partial \rho_0}{\partial t} + \frac{\partial \rho_0}{\partial a} &= -A(a; \rho_1)\rho_0 + C_D\rho_1, \quad (1.8) \\
\frac{\partial \rho_1}{\partial t} + \frac{\partial \rho_1}{\partial a} &= A(a; \rho_1)\rho_0 - [C_D + R(a; \rho_2)]\rho_1 + C_P\rho_2. \quad (1.9) \\
\frac{\partial \rho_2}{\partial t} + \frac{\partial \rho_2}{\partial a} &= R(a; \rho_2)\rho_1 - C_P\rho_2. \quad (1.10)
\end{align*}
\]

The left-hand side is the total time derivative \( d/dt = \partial/\partial t + (\partial a/\partial t)\partial/\partial a \); provided age and time are measured in the same units \( \partial a/\partial t = 1 \). The transition rates that appear above are expressed as
Fig. 1.5 Long-time radical populations of any age as a function of $C_A/C_D$ and $C_R/C_D$ and under irreversible radicalization $C_P = 0$, for the fanatic-stage model of age-differentiated interactions in [139]. In addition to utopia (no radicals present) and turmoil (radicals persist) an oscillatory regime arises at steady state for large values of $C_A/C_D$ and $C_R/C_D$. Here, periods of turmoil yield to dormant phases, defining 40-year cycles similar to what predicted by Rapoport’s wave theory of modern terrorism [68]. Parameters used in Eqs. 1.8-1.13 are $C_D = 5$, $\alpha_A = \alpha_R = 20$, $\sigma_A = \sigma_R = 10$.

\[ A(a; \rho_1) = C_A \int_{a_0}^{a_1} \mathcal{K}(a, a'; \alpha_A, \sigma_A) \rho_1(t, a') da', \quad (1.11) \]
\[ R(a; \rho_2) = C_R \int_{a_0}^{a_1} \mathcal{K}(a, a'; \alpha_R, \sigma_R) \rho_2(t, a') da', \quad (1.12) \]

where the interaction kernels $\mathcal{K}(a, a', \alpha_j, \sigma_j)$ couple populations of different ages

\[ \mathcal{K}(a, a', \alpha_j, \sigma_j) = \left[ \int_{-\tilde{a}}^{\tilde{a}} e^{-s^2/\sigma_j^2} ds \right]^{-1} \exp \left[ -\frac{(\alpha_j - a)^2 + (a - a')^2}{2\sigma_j^2} \right], \quad (1.13) \]

for $j = A, R$. Although many choices are possible, as written, the interaction kernels indicate that individuals are most susceptible to activation and radicalization at target ages $a = \alpha_j$ for $j = A, R$ and through “peer to peer” interactions with individuals of similar age but more radicalized; typical values are 20 to 30 years old. The ranges $\sigma_j$ represent the effective age overlap between populations of different ages. The overall span of radical activity is assumed be within the $[a_0, a_1]$ window, so that individuals of age $a < a_0$ are too young to influence or be influenced by their peers, while those with age $a > a_1$ may be too old or entrenched for change. The term $\tilde{a} = a_1 - a_0$ in the prefactor of Eq. 1.13 guarantees that upon integration over $[a_0, a_1]$ the kernels are independent of $\sigma_j$. Finally, boundary conditions are set as
The model defined by Eqs. 1.8–1.14 was first analyzed by integrating over the age interval \([a_0, a_1]\), yielding a compartment model for which steady states were determined. Later the full age-dependent model was analyzed to draw comparisons. For both versions of the model, age-independent and age-dependent, three steady states arise: utopia (where society is solely comprised of non-radicals, \(\rho_*^0 \neq 0, \rho_*^1 = \rho_*^2 = 0\)), a dormant state (with finite populations of non-radicals and activists \(\rho_*^0 \neq 0, \rho_*^1 \neq 0, \rho_*^2 = 0\)) and turmoil (where radicals are also present \(\rho_*^0 \neq 0, \rho_*^1 \neq 0, \rho_*^2 \neq 0\)). However, the respective basins of attraction and the associated populations differed greatly in the two formulations. For example age structure enhances radicalization in certain parameter regimes, effectively increasing the push through the \(i = 0, 1, 2\) hierarchy. In other cases, early age radicalization may drain the activist pool and thwart further radicalization. The latter scenario represents the radical ideology spreading too quickly among a few who become isolated from the rest of society, and who are not able to effectively recruit more adherents through peer pressure at the intermediate, activist level.

Most notably age-independent radicalization models do not display limit cycles whereas, the model in Eqs. 1.8–1.14 allows for cyclic behavior to arise with alternating waves of more or less radicalized individuals over the course of several generations. This result aligns with political science paradigms first presented by Rapoport and discussed in the introduction whereby extremist tendencies rise and fall over time like waves or ripples [68, 146, 147, 148]. Eqs. 1.8–1.14 provide possible mechanisms to explain wave-like behavior; in particular cyclic solutions arise only when when radicalization is aggressive (\(C_R\) is large compared to \(C_D\)) and irreversible (\(C_P = 0\)). Furthermore, realistic parameter choices result in a typical period of sustained radical population of 40 years, in agreement with Rapoport’s theory that extremists radicalize a generation of individuals, and when their influence fades due to aging, the cycle of terrorism comes to an end [149]. Simulation outcomes are displayed in Figure 1.5, showing long time radical populations of any age as a function of \(C_A/C_D\) and \(C_B/C_D\) under irreversible radicalization \(C_P = 0\). Small \(C_A/C_D\) leads to utopia; turmoil prevails at small \(C_R/C_D\). For large \(C_A/C_D\) and \(C_R/C_D\), society oscillates between utopia and turmoil, defining 40-year cycles.

Before considering more structured descriptions of the radicalization process, it must be noted that the models illustrated above, as well as most sociological theories, consider a linear pathway to radicalization and assume that the adoption of extreme beliefs is a necessary condition for the execution of violent acts. Profiles of actual terrorists show that this is not always the case [65, 150], and that some individuals holding very radical views have opted to pursue peaceful avenues to manifest their discontent. Furthermore, social movement theory suggests that terrorism is but a small part of a much wider array of potential actions arising from protest movements and countercultures [151, 152]. It is thus likely that the radicalization pathway schematically represented in Fig. 1.2 can branch out towards multiple end points, of which only a subset include violence.
Fig. 1.6 Radicalization processes on a square lattice. At $t = 0$ opinions $\phi$ are heterogeneously distributed and one radical element $\rho \neq 0$ is present at the center of the lattice. Parameter choices lead to three different possible scenarios. In panel (a) under high tolerance for different opinions, views are still heterogeneous but no radicals emerge; this is the perpetual calm scenario. In panel (b) tolerance is at intermediate values: the initial radical is able to nucleate a cohort of other extremists that at $t \to \infty$ will cover the entire lattice; this is a seeded nucleation event. Finally, in panel (c) individuals are intolerant to different views and clusters of radicals emerges throughout the lattice; this is the spontaneous radicalization case. Under perpetual calm, the radical seed is unable to radicalize anyone else; under seeded radicalization, a radical population spreads out from the initial radical through nearest-neighbor interactions. Under spontaneous radicalization, individuals radicalize even in the absence of direct contact with other radicals. Figure taken from Ref. [161]

1.4 Lattice and network models

About fifty years ago, experiments on collective decision making revealed an unexpected sociological phenomenon, termed group polarization. When faced with diverse opinion or judgement possibilities, a medium-sized group of interacting individuals did not settle on a moderate, intermediate view. Rather, interactions lead the group to find consensus on an extreme position, the one supported by its most polarized members; such outcome was more pronounced if the topic was complex and nuanced [153, 154]. Several theories were offered to explain this discovery including diffusion of responsibility, persuasion, familiarization, and cultural values [155, 156, 157]. Inspired by this work, Serge Galam and Sergio Moscovici introduced one of the first lattice type models applied to social decision making with the aim of describing group polarization [158]. Their model is reminiscent of a classical Ising model for spontaneous magnetization, where entropy, couplings, temperature are given sociological interpretations. Individuals are assigned a spin-type opinion value of $\pm 1$ and interact on a square lattice in the presence of site-dependent social...
fields $S_i$ that exert a non-uniform pressure. Neighboring sites may thus intrinsically favor different spin values, but the spin-spin couplings may drive them towards a consensus. Beyond their nearest neighbors, individuals can also interact with groups of others; the size of these groups is fixed by a given parameter, and these interactions are included in the social field. The authors find that smaller interacting groups, with less vigorous spin-spin couplings, or that allow for less dissent (the low temperature scenario) tend to polarize or “break the symmetry”, whereas larger or more tolerant groups (the high temperature scenario) yield a more moderate consensus. Not only was this work one of the first to study the shaping of extreme opinions on a square lattice, the results obtained also showed the importance of the underlying fabric of social contact (couplings and interaction group size) in determining the spread and persistence of extreme opinions. This is especially true of radical behavior where small fringe groups, or even a single individual, are able to influence large segments of society depending on delivery methods and persuasion levels, but also on their connectivity. In order to go beyond compartment models thus, one must include for structured interactions between people or communities.

Several extensions of the above work to model radicalization on a square lattice were proposed [159, 160, 161]. Typically, agents interact with their neighbors and to go through a progression of stages that take them through various levels of extremism. In Ref. [161] opinions define a continuous variable $\phi$ between $\pm 1$, allowing for neutral ($\phi \sim 0$) and polarized ($\phi = \pm 1$) views to coexist. An individual with an extreme opinion is not necessarily a radical, since he or she may harbor a polarizing view while still being tolerant of the opinion of others. Radicalization $\rho$ is here assumed to depend on how an agent responds to views that diverge from its own. Each individual on the lattice is thus subject to a tension that depends on the difference between its opinion and that of its nearest neighbors; reactions are modulated by a tolerance level, so that under large enough tension and low tolerance, individuals become radicalized. Simulation outcomes are shown in Figure 1.6. Under high tolerance a non-polarizing consensus is reached and no radicals exist, akin to the emergence of an average opinion in compartment models. This is the perpetual calm scenario of Fig. 1.6(a). When individuals are intolerant to opinion heterogeneity, radicals may spontaneously emerge, as depicted in the spontaneous radicalization panel in Fig. 1.6(c). Finally, under intermediate levels of tolerance, only if the lattice has been originally nucleated with a radical element will further radicals appear, once more underlying the importance of initial conditions. This is the seeded radicalization case of Fig. 1.6(b).

Although not without controversy, it is believed that typical social interactions such as the World Wide Web define scale-free networks, where the probability distribution $f(k)$ for a node to have $k$ links to others follows a power law $f(k) \sim k^{-\gamma}$ [162, 163, 164, 165, 166]. Here, $\gamma$ measures how well-connected the network is; typical ranges are $2 \leq \gamma \leq 3$. In a scale-free network a few nodes have many connections, the majority have only a few, and there is no sharp boundary between these two possibilities. Scale-free networks have been used to study the spread of ideologies, for example through voter models [167, 168]. Here each node represents an individual that carries a 0 or 1 opinion that adjusts depending on the opinion of the
nodes it is connected to. These interactions often lead to consensus, where a single opinion permeates the entire graph. To account for scenarios where conflict arises, such as in political debate, “zealots” are introduced similarly as to what discussed in Eqs. 1.5-1.6. These stubborn players never change their attitudes and allow for distinct opinions to persist [169, 170, 171]. Other types of networks to study how cascades of extreme views can percolate throughout society have also been considered [172, 173, 174]. In particular, small-world network models of interacting, stubborn nodes have also been shown to allow for radical views to persist; demographic surveys conducted on religious perspectives and on the state of the economy in different countries are consistent with model outcomes [174].

The radicalization progression first proposed by [115] and modeled by Eqs.1.1–1.4 was adapted to two scale-free networks with $\gamma = 3$. The two cases considered are a hierarchical network with directed connections, and a symmetric network [175, 176]. Both are built by starting with four completely connected nodes of extremists $F$ to which newcomers are progressively added; the original nodes influence the attitudes of others but can also spontaneously return to the general population $G$. The level of extremism of each added node depends on the attitudes held by its neighbors mimicking a $G \rightarrow S \rightarrow E \rightarrow F$ progression as originally proposed in [115]. In the hierarchical network the attitudes of recently added nodes are influenced by those with more seniority, but the opposite is not true resulting in a history dependent interactivity that represents groups with a rigid command-like structure. The symmetric network is more flexible so that all nodes mutually influence each other, regardless of when they joined. The question is: to what degree will the original four fanatics influence incoming nodes? What was found is that at steady state the hierarchical network is populated by a uniform non-radical population made of $G$ individuals, although cohorts of low or moderate extremists, the $S$ and $E$ groups, can spread to large numbers before vanishing. In this case, the four original fanatics set an avalanche in motion that can persist for a long time, but that eventually dissipates. This result is compatible with findings from the compartment model [115] where a sufficiently large de-radicalization rate $\gamma_1$ yields a non-radical society at steady state. However, under the same initial conditions and for a subset of $\gamma_1$ parameters the symmetric network yields a much different scenario: at steady state comparable numbers of $G, S, E, F$ nodes arise, showing that the mutual reinforcement of attitudes, which is not possible in the hierarchical network, greatly increases the spread and persistence of fanatic ideals. Behaviors on the square lattice and comparisons with the compartment model are also discussed [175]. These results confirm that connectivity structures and hierarchies can greatly influence the spread of extreme attitudes, as already shown for other epidemic phenomena evolving on networks of different types [164].

Networks used to study the evolution of diseases [177, 178] were later adapted to to describe the spread of radical religious ideologies [179]. Here, $N$ nodes are distributed in space at random so that each pair is separated by a socio-spatial distance $d$, which represents affinity, friendship or family closeness. The distance $d$ also determines whether nodes are linked or not: connections are established probabilistically drawing from a half-gaussian distribution $P(d;D)$ of width $D$ so that the
average number of connections per node is fixed at given \( \langle n \rangle \) and so that the average distance on the connected network is proportional to \( D \), \( \langle d \rangle \sim D \). When three nodes make a three-way connection to form a triangle, they are considered “clustered”; the cluster coefficient \( C \) of a network is defined as the average number of such triangles per node, over the maximal number of possible triangles that can be constructed through it. As shown in Figure 1.7(a), small \( D \) yields tight clusters of nodes with shorter links among them (large \( C \), small \( \langle d \rangle \)), large \( D \) results in more global connections with relationships established across larger distances (small \( C \), large \( \langle d \rangle \)) [177]. Once constructed, a radicalization process akin, but not identical, to the \( G \to S \to E \to F \) sequence introduced in Eqs. 1.1–1.4 is imposed on the network. One of the main differences with respect to the original compartment model in [115], is that here fanatics are further divided in two classes: foot-soldiers \( R_S \) and leaders \( R_L \) which are associated to higher mortality rates due to, for example, suicide bombings, counter-terrorism attacks or arrests. Each node \( i \) is assigned an intrinsic ideology receptiveness \( \tau_i \leq 1 \) and a radicalization state \( G, S, E, F \) so that progression along the hierarchy occurs with a probability \( 1 - (1 - \tau_i)^k \) where \( k \) is the number of radical neighbors of node \( i \). Finally, the network is quasi-static in the sense that its structure changes only due to death whereby nodes are removed and later reintroduced randomly, rearranging links. Simulations run on 1250 nodes, of which 10% at \( t = 0 \) were fanatics and the remainder part of the general, non-radical population reveal that a sufficiently large initial number of fanatics can drive the core population to radicalism before extinction.
Other modeling work considers networks where nodes carry a belief $B_i$, and are connected only if they are in close geographical proximity and their beliefs are similar. These two elements model homophily, the phenomenon of being drawn to those similar to oneself. Nodes adapt the average opinion of their neighbors depending on a spread parameter $\alpha$, and on an intrinsic node vulnerability $N_i$; the most influential neighbors are assumed to be the ones with higher degree centrality. Finally, once a critical belief is achieved, the node is assumed to have radicalized and will only interact with other radicals. Apart from scale-free networks, random and small-world networks are analyzed: vulnerability and information spread are found to be more relevant than the underlying network structure in determining the abundance of radical nodes [180].

Some scale-free network models have studied radicalization as extreme forms of protest in the environmental and animal rights movements, which at times may also lead to violent action [181]. The two movements were analyzed together since they share many similarities in their origin and ideals, yet display different characteristics in the way activists protest. Environmentalists gravitate towards demonstrative or confrontational actions, with radicals opting for minor attacks on property; animal rights extremists display a greater tendency to target people [182, 183]. In the network model introduced in [181] nodes are assumed to carry a level of criminal propensity and a susceptibility for changing their morality and self-control; these levels evolve depending on social contacts with other nodes. Empirical data collected by the authors suggests that pre-radicalized animal right extremists tend to create sub-groups who attract more of the same; peaceful protesters instead remain such even after joining an animal rights movement. There is no equivalent correlation for environmental protesters. Activist nodes in animal right networks were thus assumed to preferentially link to nodes with the same criminality levels; while environmental activists forged connections with nodes of the same morality level, mimicking homophily [184]. Among the results is that initial conditions greatly affect the outcomes of activist campaigns, and that given the same initial conditions and network configuration, the final number of criminally-minded agents in animal rights movements is much higher than in the environmental case.

Two historical and purportedly scale-free networks were also analyzed: the Cathar heresy in 13th century France – the first organized challenge ever faced by the Roman Catholic Church – and the Reformation in England in the mid 16th century that led to large scale conversions to Protestantism [185, 186]. In the first case the network was suppressed, in the latter it successfully spread throughout England. Although limited data was available, key players serving as network hubs were identified in both cases [187]. By comparing the two historical outcomes, the authors concluded that in order to stop extreme movements from permeating society, it may it be more effective to isolate key players on the network by neutralizing their close contacts, rather than risk creating cascade effects through active persecution, or even martyrdom. The authors argue that the lessons learnt may be applicable to our current times.

How followers are recruited or spontaneously join online extremist networks was the subject of a detailed analysis performed on VKontakte [188]. This is a Russian-
Fig. 1.8 Dynamics of ten pro-ISIS aggregate groups on the social network VKontakte. Horizontal bars show the lifespan of each of these groups; the process of appearing and disappearing is modeled as a coagulation-fragmentation process via Eqs. 1.15 and 1.16. The inset shows a portion of an actual aggregate-follower network on January 1\textsuperscript{st} 2015 with individual followers (blue nodes) linking to pro-ISIS aggregates (red nodes). Taken from Ref. [188].

based online social networking service, with more than 350 million subscribers, heavily used by ISIS as a means to spread its ideology in Chechen Russia. Roughly 196 pro-ISIS self-organized online aggregates (the equivalent of Facebook groups) were identified over an eight month period in 2015, attracting more than one hundred thousand adherents. The growth and decline of the size of these aggregates was tracked over time revealing merging and shutdowns, possibly due to government or regulator intervention as shown in Fig. 1.8. The dynamics was then modeled as a coagulation and fragmentation process [189] as follows

\[
\frac{\partial n_s}{\partial t} = \frac{v_{\text{coal}}}{N^2} \sum_{k=1}^{s-1} k(s-k)n_k n_{s-k} - \frac{2v_{\text{coal}}}{N^2} s n_s \sum_{k=1}^{\infty} k n_k - \frac{v_{\text{frag}}}{N} s n_s, \tag{1.15}
\]

\[
\frac{\partial n_1}{\partial t} = -\frac{2v_{\text{coal}}}{N^2} n_1 \sum_{k=1}^{\infty} k n_k + \frac{v_{\text{frag}}}{N} \sum_{k=2}^{\infty} k^2 n_k. \tag{1.16}
\]

Here \( n_s(t) \) is the size of aggregate \( s \) so that the total number of individuals is \( N = \sum_s n_s(t) \). The three terms on the right hand side of Eq. 1.15 refer respectively to: coagulation/merger between an aggregate of size \( k \) and \( s-k \) to form a larger one of size \( s \); the loss of an aggregate of size \( s \) that merged with another one; the fragmentation or loss of an aggregate due to its dissolution. Unaffiliated individuals belong to the \( s=1 \) aggregate. Rates of coalescence and fragmentation are given by \( v_{\text{coal}} \) and \( v_{\text{frag}} \) respectively. Eq. 1.15 is complemented by Eq. 1.16 which models the dynamics for an aggregate of size \( s=1 \) and where no coalescence is possible, since the minimum size of a coalescence process is \( s=2 \). The two terms on the
right hand side represent respectively the merging of an $s = 1$ aggregate with any other, and an increase in the number of $s = 1$ aggregates due to fragmentation of an existing one. Using generating functions [190] it can be shown that at equilibrium the size distribution of aggregates $n_s^*$ scales according to

$$n_s^* \sim \left( \frac{\nu_{\text{coal}}^{s-1}(\nu_{\text{coal}} + \nu_{\text{frag}})^s}{(2\nu_{\text{coal}} + \nu_{\text{frag}})^{2s-1}} \right) s^{-\frac{5}{2}}.$$

In the limit of large $s$ a power-law distribution emerges $n^*_s \sim s^{-\gamma}$ where $\gamma = 2.5$, in good agreement with the $\gamma = 2.33$ seen in observations. Eqs. 1.15 and 1.16 also allow to explore possible avenues of anti-ISIS online intervention. For example, it is shown that an effective strategy is for anti-terrorism agencies to focus on the shutting down of smaller groups, before they become too potent; indeed, if the dissolution rate $\nu_{\text{frag}}$ is too slow, several clusters will rapidly coalesce into one super-aggregate. Finally, the same authors used a similar coagulation-fragmentation framework to describe the likelihood that a terrorist attack will result in a given number of victims. Richardson’s Law is a well known paradigm in political conflict, according to which the distribution of casualties due to violence follows a power law, so that the probability $p(x)$ of an event with $x$ deaths scales as $p(x) \sim x^{-\gamma}$. The assumption made in [191, 192, 193] is that an insurgent movement is made of various cells of extremists, each carrying an attack strength $s_i$ that represents the potential number of casualties it will inflict. Cells are continuously rearranging their attack strengths through coalescence and fragmentation, as described by Eqs. 1.15-1.16 so that the probability distribution for attack strength $n_s^* \sim s^{-\gamma}$ in Eq. 1.17 can be viewed as a proxy for the number of casualties $p(x) \sim x^{-\gamma}$. A convincing illustration of this parallel is that the scaling of the number of victims due to insurgent attacks in Iraq, Colombia and Afghanistan follows a power-law distribution, with exponents consistent with the $\gamma \sim 2.5$ estimate predicted by Eqs. 1.15-1.16.

But how exactly is a network of operational terrorists structured? How does it change in time, due to internal dynamics, changes to the socio-political environment, the need for secrecy, technological advances, and in response to counter-terror interventions [194]? Any progress in answering these questions may lead to a better assessment of terrorist hierarchies, and in developing policies and practices to best detect and disrupt them [195]. Terrorist groups function underground and their secretive nature makes data collection challenging; to date, information has mostly been gathered via open-source texts from the media and publicly available legal transcripts [196]. Initial efforts to understand terrorist organizations sprung in the early eighties, to catalogue insurgent activities in Palestine, by the Provisional Irish Republican Army (PIRA), and covert operations by the KBG [197]. The September 11th terrorist attacks led to greater urgency in this direction; as much as possible, data collection, social network and text analysis were used to identify emergent leaders, spheres of influence, hubs, graph structures of individual cells and of extremists frequenting web forums [4, 198, 199]. Several computer-supported techniques combining network text analysis and methods to classify organizational systems or so-
Fig. 1.9 An agent-by-agent Middle-East terrorist network visualized through the Meta-Matrix Text Analysis [205]. Media outlets searched to create this graph includes major newspapers, magazines, journals, trial transcripts, book excerpts, scientific articles on Al-Qaeda in Iraq from 1977 to 2004. The nodes on the left are not directly linked to other actors. The isolated, circular sub-graph on the left side of the inner circle that is not connected to other agents represents individuals who were charged with the Khobar Tower Bombing in Saudi Arabia in 1994. Relevant nodes were identified: one carries the highest cognitive demand (0.0915), degree centrality (0.0915) and betweenness centrality (0.0463); a distinct node has the highest clique count (12.0), and yet another distinct node has the highest number of simmelian ties (0.0282). Taken from Ref. [202].

The nodes on the left are not directly linked to other actors. The isolated, circular sub-graph on the left side of the inner circle that is not connected to other agents represents individuals who were charged with the Khobar Tower Bombing in Saudi Arabia in 1994. Relevant nodes were identified: one carries the highest cognitive demand (0.0915), degree centrality (0.0915) and betweenness centrality (0.0463); a distinct node has the highest clique count (12.0), and yet another distinct node has the highest number of simmelian ties (0.0282). Taken from Ref. [202].

What emerges is that terrorist networks tend to display a highly centralized organization, but counter-terror pressures compel them to modify their structures and actions, forcing them to find compromise between security and communication.

cial network analysis [200] were introduced, such as the Meta-Matrix Text Analysis [201, 202], the Dynamic Network Analysis [203, 204], the Computational Analysis of Social and Organizational Systems [205] the Counter-Terror Social Network Analysis and Intent Recognition [206]. These packages evolved over time to allow for visual, sophisticated representations of connections of actors, locations, events, tasks, knowledge, resources, allowing to study not just command hierarchies, but also relationships based on economics or training. Some software also allow for the removal of actors to forecast possible consequences to the terrorist structure [204] after their capture or death. These features can be especially useful for diplomatic or military considerations. One example of a network reconstructed via the Meta-Matrix Text Analysis is shown in Fig. 1.9.
Typically, leadership roles are re-adjusted and hierarchical organizations become more decentralized so that the network may survive a negative environment. This trend is observed in many networks, including in al-Muhajiroun, a British group that advocates the establishment of an Islamic state in the country: once its leader left Britain, followers adapted their operations, created spin-off groups, kept recruiting new members, and shifted towards online communication. A similar trend was seen in the the militant Islamic group led by Noordin Top in Indonesia between 2000-2010: as authorities tried to disrupt the organization, its topography changed in terms of centralization and density. Similarly, Al-Qaeda abandoned its initial corporate-like command structure in favor of a so-called leaderless jihad due to increasing scrutiny after the September 11th terrorist attacks. These organizational changes are also confirmed by the text analysis methods described above, where over the course of a decade, the density and communication levels within Al-Qaeda cells were seen to decrease. An evolutionary process to describe Al-Qaeda’s path from a tightly coupled organization, to a coupled network, to loosely coupled movement is illustrated in. Furthermore, in its post September 11th incarnation, Al-Qaeda has been described as a “dune-like” organization whereby its affiliates are encouraged to operate independently while still part of its broader network. Among the many drawbacks to this tactic is that rebel groups may splinter and antagonize the once centralized command. This is the genesis of present-day ISIS that began as part of the umbrella of organizations under Al-Qaeda’s influence, but that rapidly emerged as one of its strongest contenders.

Since the few, but well connected, nodes typical of scale-free networks may jeopardize the operational security of terrorist cells, it is important to understand which organizational structures best serve covert organizations. The tradeoff between operational efficiency and secrecy in terrorist networks was analyzed in where various models of communication on connected graphs were compared and contrasted under different detection risk scenarios. In this work, each node $i$ of a graph is characterized by a degree $d_i$. The total number of nodes is $n$ and the total number of links is $m$. First, a communication measure $I(g)$ is defined based on the connectivity of the graph

$$I(g) = \frac{n(n-1)}{\sum_{ij} \ell_{ij}(g)},$$

(1.18)

where $\ell_{ij}$ is the shortest distance between nodes $i$ and $j$, and $n(n-1)/2$ is the number of pairs in a network of $n$ nodes. The quantity $I(g)$ can be thought of as the normalized reciprocal of the total distance and as a result $0 \leq I(g) \leq 1$. Since $\ell_{ij} \rightarrow \infty$ for unconnected nodes, $I(g) \rightarrow 0$ for a disjoined graph; $I(g) = 1$ on a fully connected graph. Two factors are assumed to contribute to the security of node $i$: the exposure probability of being identified as member of the network, $\alpha_i$, and the fraction of the network that remains unexposed after a given node has been detected, $u_i$. The overall likelihood of the network remaining unexposed is thus given by
An optimal graph is one that maximizes the product of the information $I(g)$ and secrecy $S(g)$ measures, $\mu(g) = S(g)I(g)$. The first scenario analyzed is that of a uniform exposure probability $\alpha_i = 1/n$ and where if node $i$ is identified as part of the network, all others connected to it become exposed with a link detection probability $p$, so that $u_i = (n - pd_i - 1)/n$. Here, $p = 1$ represents the case where all links are exposed; $p = 0$ indicates an impenetrable network where the identifica-
tion of one node still allows for the other $n - 1$ to remain unknown to counter-terror agents. The other scenario is that of node detection being proportional to its centrality in the network, $\alpha_i = (d_i + 1)/(2m + n)$ and where a detected node leads to all its connections being exposed $u_i = (n - d_i - 1)/n$. Game theoretic Nash bargaining is then used to determine which $g$ structure leads to maximal $\mu(g)$ for small values of $n$ [217, 218]; simulations are used in the case of larger $n$, where $\mu(g)$ is explicitly evaluated over multiple configurations and the maximum value recorded. In the first case, if the link detection probability is high, $p \to 1$, the Nash bargaining solution is a star graph. As the link detection probability $p$ is decreased, a transition is observed and a complete graph becomes best. In the second case, cellular networks such the ones shown in Figs. 1.10(a)–(c) are optimal. As a result, optimal covert networks should incorporate all-to-all communication when detection risks are low; a star configuration when communications can be intercepted but nodes are equivalent; reinforced rings, wheels and other cellular structures if centrality is important. Other choices of $I(g)$ may lead to different optimal graphs; for example in Fig. 1.10(d) best structures are shown if the communication measure is set at $I_2(g) = 1/D(g)$ where $D(g)$ is the diameter of the graph defined as the maximal distance among all nodes, $D(g) = \max_{i,j \in g} \ell_{ij}$. Among the many variants of this work, one of the most notable is the inclusion of node heterogeneity where a specific interaction between a pair of nodes may be riskier than others. In this case network structure is optimized by having the pair in question being the least connected to the remainder of the network. This theoretical framework was later applied to the 2002 Bali attacks by Jemaah Islamiya, an Indonesian Al-Qaeda affiliated Islamist group, for which prior knowledge of operational and command nodes was known [199]. Links were assigned heterogeneous weights representing different risk levels. The optimal networks for $n = 25$ and $n = 40$ consist of cellular structures around a centralized individual, less dense than those observed in Fig. 1.10(c) since the high risk link carries limited connectivity to the remainder of the network [216, 219].

The structure of actual operative cells that attacked five South East Asia locations and Madrid, Spain was analyzed in Ref. [200]. What was found is that all terrorist networks became increasingly dense and cohesive as they prepared for executing the above attacks; these networks did not display scale-free characteristics, nor did they seem to reorganize much to optimize communication secrecy. This has led to the interesting theory that as they ready for action, and especially if they operate under a sense of security, terrorist groups will not devote resources to optimize their internal structure, as it would distract from their main objective, which is to attack. It is only the threat of detection and the need for covertness that catalyzes changes to existing patterns, since links that were forged over time, not necessarily in an organic way and connecting individuals with different ideals, skills and personalities, require costly, time-consuming intervention to rearrange. As a corollary, it is suggested that a possible counter-terrorism strategy could be to foster a false sense of security, allowing for connections to densify over time so that once intervention occurs the maximal number of members of the network will be exposed. The same authors propose that rather than eliminating hubs of well-connected individuals hoping that the network will collapse, a better strategy in the absence of
a scale-free structure, or in the presence of networks that regenerate themselves, is to employ enough resources to target the entire terrorist cell. The case of terrorists organized in a clique of $N$ nodes, that is a graph where every two nodes are directly connected, has been analyzed in [220]. Because of the structure of a clique, the capture of any of its members would expose all. If detection probability is set at $p$ per link, percolation theory allows to estimate that cliques of size $N \geq 1/2p^2$ will be detected with high probability. This non-linear outcome suggests that any investment to increase surveillance to $p$ would translate to capture rewards of order $p^2$; seen from a different perspective this result implies that to avoid detections, as $p$ is increased, cliques must shrink by $p^2$ members. Other work addresses the question of how hierarchical terrorist groups would function when crucial members are captured or killed [221] and advocates for searching network cut-sets, that is nodes whose removal disconnects leaders from followers.

The organizational dynamics of the Global Salafi Jihad (GSJ) terrorist group, a Muslim revivalist movement, was studied between 1989 and 2003 using open source information [222]. Results from this test case were consistent with the dynamical changes described above: the network was observed to be rather random at the onset, but acquired a scale-free topology as new members joined while others left, due to the pressure from counter-terrorism operations. Average degree, degree distribution and network diameter, defined as the average length of the shortest path between any pair of nodes, were used as statistical measures. Furthermore, vulnerability to random failures, targeted attacks, and counter-terrorism efforts were all analyzed by simulating node removal; the loose network was found to be resilient to all disruption efforts.

The Stochastic Opponent Modeling Agent (SOMA) was originally developed to understand how social, economics or religious communities behave; later it was specifically applied to terrorist organizations [223, 224, 225]. Datasets tracking rebellion and protest movements worldwide are mined to generate rules of behavior for known violent ethno-political or ethno-nationalist groups [226, 227]. These rules are associated to given actions such as kidnapping or orchestrating transnational attacks, and to various environmental characteristics, such as receiving financial or military support from a foreign country. From the data, one can then reconstruct likelihoods for a given organization to execute a given action under given conditions. This is valuable since the number of action-condition combinations is very large and a human analyst could easily miss an interesting hypothesis. Hezbollah for example was tracked for 23 years and one finding is that when involved in local inter-organizational conflicts, it is less likely to engage in transnational violence [228]. Hamas was also studied [229] and appears to especially engage in violent activities when providing social services to Palestinians; its kidnapping rate of Israeli citizens also increases at times of heightened conflict with Fatah, its major rival in the area. Other contexts in which SOMA-rules were applied include understanding the behavior of actors in the Afghan drug economy [230] and of separatist movements in the Jammu and Kashmir provinces of India and Pakistan [231]. These findings do not explain why certain correlations are observed, however they are still useful as they pose new questions, offer new hypothesis and food for thought.
The topological role of women in terrorist organizations is explored in [232]; data mining performed on the social media outlet VKontakte and on offline records detailing the 1970-1988 activity of the Provisional Irish Republican Army (PIRA) show that in both organizations women occupy nodes with superior network connectivity, guaranteeing robustness and survival to the network itself [232]. Betweenness centrality is a good indicator to measure how influential a given node is in a network; it is defined as the fraction of shortest paths connecting any two nodes that pass through a given node. In a network of covert extremists, understanding shortest path features is especially relevant, since adding any extra step to connect players represents risk and potential cost. The analysis in [232] shows that betweenness centrality on female nodes is much higher than on male nodes. The high interconnectivity of women implies that play important roles as communicators of messages and goods; it would be interesting to include this feature into future dynamical models of terrorist networks.

1.5 Game theoretic models of terrorism

In previous sections we discussed the radicalization of individuals, and how new recruits join terrorist organizations. We also reviewed the social structures of terrorist networks that facilitate communication while evading detection, and how various cells may merge or dissipate. All these human and material resources determine the organizational capability of a terrorist organization; whether such capabilities translate into a successful execution however, depends on how extremists interface with existing deterrents. Before striking, terrorist organizations must rationally weigh their probability for success, and evaluate their expected gains against potential costs and the risk of being exposed. Attacks may thus be executed in the immediate, postponed to future times, or aborted. The process of rational decision making is best explored through the tools of game theory. Aerial hijacking and hostage taking incidents, which were commonplace in the eighties, were among the first terrorist incidents to be described via game theoretic models [233, 234, 235, 236, 237]. Despite the fanatical ideologies involved, negotiating strategies and reactions involving terrorists and counter-terrorism agents were found to be remarkably rational. This was interpreted as a consequence of both sides seeking to predict future developments in volatile situations, so as to avoid unexpected occurrences [238]. More recently, differential games have been used to model the dynamics between terrorist organizations and government authorities, in an attempt to estimate the risk of terrorist attacks and to assess prevention strategies [239, 240, 241, 243, 244, 245, 246, 247, 248, 249].

Differential terrorism games originate from dynamical compartment models [250, 251]: extremists and governmental counter-terror agencies adjust the intensity of their activities as they interact with each other and other sociopolitical entities, seeking the most favorable outcome. The strength of a terrorist organization is described by a scalar state variable $x(t) \geq 0$, which represents one or a combina-
ition of resources the organization possesses, such as number of recruits, weapons, supporters from the general public, financial capital, knowledge and access to new technologies [240, 241, 252]. The state variable $x$ evolves over time as follows [244, 246, 248, 249]

$$\frac{dx}{dt} = I(x; u) - O(x; u) = \mathcal{E}(x; u),$$

(1.20)

where $I$ and $O$ represent the influx (accumulation) and the outflux (consumption) of resources, respectively, and $\mathbf{u}(t)$ denotes an array of time-dependent control parameters that reflect strategy adjustments of both terrorists and governmental agencies. The influx term $I$ is typically assumed to be independent of external circumstances and to grow linearly at rate $r$, $I(x) = rx$ [246, 248, 249]. Some models include a logistic growth $I(x) = rx(x_{\text{max}} - x)$, where $x_{\text{max}}$ specifies the maximum possible resource [241]. A constant influx is sometimes included to “jump start” the model and usually set to a very small value [243, 245]. Further dependence on the control parameters $\mathbf{u}$ may be included. For example, an overly aggressive counter-terrorism policy may antagonize the general public, boosting public support for terrorist groups, and favor the recruitment of new members [240, 241, 243]. The most simplistic assumption is for the outflux term $O$ to depend on a two-component control parameter $\mathbf{u}(t) = (u_1(t), u_2(t))$ where $u_1(t) \geq 0$, $u_2(t) \geq 0$ represent the intensities of terrorist activities and anti-terrorism operations, respectively, so that [246, 248, 249]

$$O(x; \mathbf{u}) = h(u_1, u_2).$$

(1.21)

The function $h$ is referred to as a “harvest”; it specifies the consumption and loss of terrorist resources due to attack executions, arrests, foiled attacks, and other events. It is typically assumed that $\partial h/\partial u_i > 0$ for $i = 1, 2$ and for all $u_i$ values, so that increases in terrorist and governmental activity result in increases to the harvesting of resources. Further assumptions are made on the second-order partial derivatives of $h$: $\partial^2 h/\partial u_i^2 > 0$ indicates that as the intensity of terrorist activities increases, the consumption of resources further intensifies; $\partial^2 h/\partial u_1^2 < 0$ indicates that as the intensity of anti-terrorism activity increases, terrorist resource consumption increases at a slower rate. The underlying idea behind this assumption is that as counter-terrorism measures intensify, decreased gains are to be expected in terms of losses to terrorists. Finally, $\partial^2 h/(\partial u_i \partial u_j) \geq 0$ indicating that increased attack intensities lead to greater anti-terrorism operations, whereas increased anti-terrorism activities will increase terrorist consumption of resources. Lastly, $h$ satisfies the boundary conditions $\lim_{u_i \to 0} h(u_i, u_j) \to 0$, and $\lim_{u_i \to \infty} h(u_i, u_j) \to \infty$ for $i, j = 1, 2$. These are known as the Inada conditions [253]: there is no spending of resources if there are no terrorist attacks nor any anti-terrorism measures in place; when either terrorists or governmental agencies strongly increase their activities, the harvest of resources becomes infinitely large. Some models include natural attrition to the outflux of resources, such as departure of personnel and expiration of supplies, which limits resource growth [245]. Both terrorists and anti-terrorism authorities aim to maximize given
utility functions by varying their respective control parameters \(u_i\) for \(i = 1, 2\) under the constraints given by Eq. 1.20. The terrorist utility function is written as

\[
J_1(T) = \int_0^T e^{-\rho_1 t} \left[ \alpha x(t) + \beta u_1(t) \right] dt + e^{-\rho_1 T} [\alpha x(T) + \beta u_1(T)].
\]  

(1.22)

Eq. 1.22 implies that terrorists benefit from increasing resources \(x\) and attack intensities \(u_1\): \(\alpha > 0\) and \(\beta > 0\) define their relative weights, respectively. The exponential pre-factor indicates that resources and attack capabilities decay in time at rate \(\rho_1\). It is important to note that \(\rho_1 \leq r\) leads to unrealistic scenarios, whereby the best strategy for the terrorist group is to accumulate resources and capabilities \textit{ad infinitum} without ever attacking. Hence \(\rho_1 > r\) is always implicitly assumed. The termination time \(T > 0\) specifies the time horizon over which the optimization game is to be conducted. For example, \(T\) may be a presidential term if the terrorist organization aims to undermine the incumbent administration; incumbents may instead seek political capital through counter-terrorism measures as they strategize for the next electoral cycle [241]. If no specific timeframe is considered, \(T \to \infty\), so that the second term in Eq. 1.22 vanishes. Governmental agencies seek to maximize the loss of terrorist resources, the \(h(u_1, u_2)\) harvest, and to minimize \(x\) and \(u_1\) using as little effort \(u_2\) as possible. These assumptions leads to the following government utility function

\[
J_2(T) = \int_0^T e^{-\rho_2 t} \left[ \gamma h(u_1(t), u_2(t)) - \kappa x(t) - \sigma u_1(t) - \eta u_2(t) \right] dt + e^{-\rho_2 T} \left[ \gamma h(u_1(T), u_2(T)) - \kappa x(T) - \sigma u_1(T) - \eta u_2(T) \right],
\]  

(1.23)

where \(\gamma > 0\), \(\kappa > 0\), \(\sigma > 0\), and \(\eta > 0\) are constant weights. Just as for \(J_1(T)\), in order for realistic scenarios to emerge \(\rho_2 > r\); if this condition is not obeyed, the best strategy for governmental agencies would be to allow extremists to keep accumulating resources and attack capability, which is undesirable.

The simplest scenario is to assume that both terrorist organization and authorities do not wait for each other’s actions to establish their own tactics. The model is thus an open-loop game. To proceed, for simplicity, one can let \(T \to \infty\) so that \(J_1 = \int_0^\infty j_1(t) dt\) where \(j_1(t) = e^{-\rho_1 t} (\alpha x(t) + \beta u_1(t))\); similarly for \(J_2\). To calculate the optimal \(u(t)\) under the constraint given by Eq. 1.20 a Lagrangian function \(\mathcal{L}_1(x, \lambda_1, u, t)\) is derived by extending the \(J_1\) payoff to incorporate the constraint. This is done through a continuum of Lagrange multipliers \(\lambda_1(t)\) resulting in

\[
\mathcal{L}_1 = \int_0^T j_1(t) + \lambda_1(t)(\mathcal{E}(x; u) - \ddot{x}(t)) dt.
\]  

(1.24)

The quantity \(H_1(x, \lambda_1, u, t) = j_1(t) + \lambda_1(t)\mathcal{E}(x; u)\) is referred to as the Hamiltonian; the corresponding \(H_2(x, \lambda_2, u, t)\) can be derived for the utility \(J_2\), subject to the same constraint. Using Pontryagin’s maximum principle the optimal control \(u = (u_1, u_2)\) and the optimal trajectory \(x(t)\) can be found by minimizing the \(H_1, H_2\) Hamiltoni-
ans [254]. The resulting Nash equilibria depend on the form of the harvest function \( h(u_1, u_2) \): if \( h \) increases at a relatively steep rate as a function of \( u_2 \), the optimal strategy is for terrorists to attack less and \( u_1 \) is kept low; if counter-terrorism operations strongly affect terrorist activity, the best strategy for governmental agencies is to increase \( u_2 \).

In more realistic settings, one side observes the actions of the other and adjusts its strategy accordingly. For example, the terrorist organization may secretly observe how the government deploys its resources before choosing whether and which target to attack; similarly once counter-terrorism measures are in place, authorities may wait for terrorist organizations to strike first, and later adjust their responses. When one player is a leader and the other is a follower in a two-player game, the model is a closed-loop Stackelberg competition, for which sub-game perfect Nash equilibria may exist [255]. Such equilibria can be found by solving the optimal strategy of the follower \( u^*_F \) as a function of the leader’s \( u^*_L \); the leader determines its optimal strategy \( u^*_L \) accordingly. Note that the Hamiltonians are different in the open-loop and in the Stackelberg games.

Regardless of which player takes on the role of the leader, the Stackelberg game always results in higher payoffs to the leader than in the open-loop game. This signifies that, given the same circumstances and model parameters, a good strategy for either terrorists or anti-terrorism agencies is to be proactive and act first. When terrorists are the leaders, both groups act more cautiously than in the open-loop scenario, there are fewer terrorist attacks and fewer counter-terrorism operations, but more resourceful terrorists at the end of the game [246]. If governmental agencies are the leaders, optimal strategies on both sides depend on \( \sigma \), the weight assigned to \( u_1 \) in Eq. 1.23. This quantity can be interpreted as the damage inflicted on governmental assets per terrorist attack. For relative small \( \sigma \), \( x(t) \) becomes negative at long times, and a Nash equilibrium does not exist. For moderate \( \sigma \), when damages inflicted are modest, the equilibrium \( u_2 \) value is smaller, and \( u_1 \) larger than in the open-loop game: counter-terrorism agencies are more guarded, and this allows terrorists to accumulate more resources. Finally, for large \( \sigma \), when attacks are on a larger scale, the equilibrium \( u_2 \) is larger and the equilibrium \( u_1 \) is smaller in the Stackelberg than in the open-loop game: counter-terrorism agencies are more proactive in preventing these large scale attacks, and extremists become more cautious.

Eq. 1.20 may be extended to explicitly include public support of the government, which may grow in time, decrease due to terrorist attacks, or be boosted by successful anti-terrorism operations [248]. The two players may also play a zero-sum game, where one seeks to minimize the payoff of the other, rather than maximize its own utility [249].

Excessively aggressive anti-terrorism tactics that can alienate public opinion and increase support for terrorist organizations have also been investigated [241]. Here, \( I(x; u) \) is augmented by \( I_{ant}(x; u) \propto u_2(t)^2 \), representing unintended advantages that government efforts yield to terrorists, limiting their own intervention capability. In most scenarios, a perpetual tug emerges between the two parties: by provoking authorities to respond overaggressively, terrorist attacks may self-sustain themselves despite the large consumption of resources, and a cycle of violence emerges. This
scenario is avoided if the proper values of $\alpha, \beta$ are selected in Eq. 1.22, in particular if the terrorist group is more interested in the collection of resources (large $\alpha$), and less in the execution of attacks (low $\beta$). It is thus suggested that governmental agencies should attempt to influence terrorists to pursue political goals rather than conduct destructive acts; this shift may also emerge naturally as extremist groups mature. Historically, many former terrorist organizations did transition into legitimate political parties, and at least to some degree renounced violent methods. Well known examples include the Revolutionary Armed Forces of Colombia (FARC), the Provisional Irish Republican Army (PIRA), and at least partially, the Palestine Liberation Organization (PLO) and Hamas in the Middle East. This conclusion is similar to what observed in Ref. [131] and discussed in Section 1.2, where using a different game-theoretic mathematical framework, it was shown that the optimal strategy for a small radical faction to grow after having established itself through violence, is to gradually disengage from it and focus on indoctrination.

The effects of government authorities applying negative (“sticks”) and positive (“carrots”) counter-terrorism incentives is analyzed in [242] under the assumption that negative incentives may be detrimental causing, for example, the emergence of hatred in the local population. A related “fire or water” model was also developed whereby two different anti-terrorism strategies can affect the strength of a terrorist group $0 \leq x \leq 1$ [243]. Fire strategies to contrast $x$ are cheaper, more aggressive, and more effective but less precise; they are also more controversial and may indirectly advantage terrorist groups. Water strategies are less invasive, precise and surgically planned, and remove terrorist threats without antagonizing the general public. They are also assumed to be less effective and more costly than fire strategies. The objective of anti-terrorist authorities is to minimize terrorist strength $x$, as well as expenditures for water $u$ and fire $v$ strategies; reactions from terrorists and related costs are disregarded. As a result, $\sigma = \gamma = 0$ in 1.23 so that terrorist responses $u_1$ and the harvest $h$ are not included in the expression for the $J_2$ government utility function. Fig. 1.11(a) shows that as $x$ increases, the optimal intensity of fire strategies $v^*$ also increases, signifying that stronger terrorist groups require more vigorous intervention. The optimal water intervention $u^*$ stays instead relatively constant. For $x$ large enough however, the dynamics of the system leads terrorist strength towards a finite non zero Nash equilibrium $x \to x_E$. Since payoff functions aim to minimize terrorist activities but also government expenses, eradication of the extremist groups does not necessarily yield the optimal utility. Fire tactics are thus best suited when the initial strength of the terrorist organization is larger than its equilibrium, $x > x_E$, since they will lead to a partial weakening of the extremist group. A second equilibrium $x_D$ emerges for very small values of $x$, with $x_D \ll x_E$ as shown in Fig. 1.11(b). Here, for $x \leq x_D$, the corresponding optimal water strategy $u^*$ drives $x \to 0$, completely eliminating terrorist activities. A discontinuity in $u^*$ arises at $x = x_D$ with two distinct values for optimal payoff: the higher one, for $x \to x^+_D$ leads to $x \to 0$; the lower one for $x \to x^-_D$ results in $x \to x_E$. Also note that for $x$ smaller than a given threshold, $x < x_S$, the optimal fire strategy $v^* \to 0$, suggesting when terrorists are not very resourceful, water strategies are the sole optimal way to completely eliminate terrorism.
Citizen approval of anti-terrorism operations is introduced to the model defined in [245] through an additional state variable $y(t)$ that represents support from the general public. This quantity increases as terrorist organizations grow their resources $x$, but decreases under aggressive counter-terrorism operations

$$\frac{dy}{dt} = r_y x - \omega u_2^2 + \zeta (y_{\text{max}} - y), \quad (1.25)$$

where $r_y$ is rate of public support, $\omega$ describes antagonistic effects due to overaggressive counter-terrorism initiatives, and $\zeta$ specifies the rate of $y$ converging to a maximum $y_{\text{max}}$ under neutral conditions. Higher public approval is also modeled to increase the effectiveness of counter-terrorism operations, as it may lead to larger budgetary means to combat extremism, or even in active participation in seeking fugitives or reporting suspicious activities. A $y$-dependence is thus introduced to the harvest function $h(y, u_1, u_2)$, such that $\frac{\partial h}{\partial y} > 0$ for $x, u_1, u_2 > 0$. Multiple equilibria are found depending on parameter choices. In one case, the optimal scenario is the counter-terrorism authority enjoying public support and effectively suppressing terrorism; in the opposite case negative public opinion hinders anti-terrorism efforts and allows extremists to thrive. Finally, a specific set of parameters yields bistable solutions depending on initial conditions: here, strong public support at the onset of counter-terrorism operations is shown to be crucial for terrorist activity to be effectively contained. Sensitive regions in phase space are also identified whereby small parameter perturbations may strongly affect the final equilibrium configu-
ration: such volatility may represent both a risk and an opportunity for counter-terrorism agencies as they develop intervention strategies.

A three-way interaction model that includes counter-terrorism efforts, public opinion, and a community that harbors terrorists is presented in [256]. The harboring community is assumed to contribute to anti-terrorism efforts, as long as the scale of these operations is limited. However, once governmental involvement exceeds a given threshold, the harboring community becomes antagonistic: costs are increased and the effectiveness of counter-terrorism efforts are reduced. Public opinion on the other hand is assumed to be supportive of anti-terrorism programs that decrease attack risks; these programs may be costlier. As a result, counter-terrorism agencies must weigh whether to increase the $u_2$ intensity of their operations to garner public support, or risk alienating the harboring community, increasing related costs. It is found that by applying legal constraints to counter-terror efforts, effectively imposing bounds to the acceptable values of $u_2$, public opinion and the harboring community may both be steered towards being more supportive of governmental intervention. Similar issues were explored in Stackelberg games where limits were posed on interrogation methods and other tactics to extract information from detainees [257].

Game theoretic models have also been applied to study terrorism from different perspectives. For example, [258, 259] optimized patrol schedules to protect crucial government assets from terrorist attacks, such as oil pipelines and chemical plants, under the constraint of limited resources; in the same vein [323] consider a leader-follower game where the state (the leader) installs facilities and a terrorist group (the follower) attacks metropolitan areas. The game is played out so that losses due to terrorist attacks are minimized by the strategic placement of these facilities. How to best allocate resources to defend sites at risk is also discussed in [261] where the distinction is made between damage due to chance events and targeted action by ill-intentioned parties. In the former case, the optimal policy is to invest resources on target high-priority sites, in the latter to spread resources and protect even the most vulnerable areas. Other defense strategies are presented in [262, 263, 264]. The work of [265] analyzes the synergy between terrorists and the active or passive support of one or more state sponsors as they form a coalition to attack a third party target. Here, the interplay between players changes in time: the state sponsor may use the terrorist group as a way of covertly promoting hostile acts towards the third party, but the latter may increasingly become aware of manipulations and retaliate. Brinkmanship game theoretical models have also been proposed: here one or both parties push dangerous actions, on the brink of disaster, to get the most advantageous outcome. The challenge is to issue threats that must be sufficiently unpleasant to deter terrorists, but not so repugnant as to not be carried out; these models thus carry a credibility constraint [266]. The effects of counter-terrorism policies have been studied in two strategic cases [267]: when government intervention is defensive, and a corollary to terrorist action, and when it is a proactive substitute for it. Negative responses may arise in both cases: erosion of terrorist support, if attacks cause too much damage; backlash against the government, if its response is too strong. It is found that large-scale attacks will occur if government reactions
produce a strong enough backlash. Game-theoretic network centrality, which identifies the most influential nodes within a cooperative game, have proven to be useful in identifying key members of terrorist networks [268]. Search games have also been proposed as methods to hunt key radical operatives: terrorists attempt to maximize the time to search terrorist networks, while government tries to minimize it [247]. The scenario of former terrorist organizations seeking a peaceful resolution to ongoing conflict has also been analyzed [269]. Here, a game of learning is used to describe negotiations between the two parties, whereby governmental agencies infer the willingness and the ability of terrorists to commit to the peace process by observing their actions. Evaluations are then used by the anti-terrorism agency to decide whether to keep pursuing negotiations or abandon them. Finally, since transnational terrorist activities can affect policy making in different countries, sequential games have been introduced involving multiple governmental players so that a country may choose between a preemptive or a defensive counter-terrorism strategy in response to the outcome of another country’s strategy. The two may also rationally evaluate whether it is beneficial to forge an anti-terrorism coalition or engage individually [270, 271].

### 1.6 Terrorist events as self-exciting processes

One major question in the study of terrorism is whether the risk of future attacks can be quantified, drawing on information gathered from past events and using mathematical modeling, risk management or machine learning methods [72, 78, 272, 273, 274, 275, 276]. It is well known that terrorist activity is not entirely random and that a prior attack can temporarily increase the likelihood of another occurrence in its geographical vicinity [277, 278]. First attempts in this direction are found in early studies of terrorism where various time-series analysis were used to study the frequency of terrorist attacks [279, 280, 281, 282, 283, 284, 285], and to explain both the temporal clustering of events and the long quiescent periods in between [286, 287, 288, 289].

Mathematically the phenomenon of past events acting as catalysts for future ones can be described as a self-exciting temporal point process, also known as the Hawkes process [290, 291] which was first introduced to study contagious phenomena and to predict earthquake aftershocks [292, 293]. Given that social behaviors are often of a contagious nature and transmit through human interactions, a wide range of social activities have been studied as self-exciting phenomena. These include committing crimes [294, 295], the sparking of riots [296], the erupting of political disorder and violence [297], the occurrence of political turnovers [298] and military coups [299], the clustering of suicides [300], and even the exchange of email messages [301]. Within the context of terrorism, self-excitation markers have been found in airplane hijacking incidents during the eighties [302], and in insurgent activities over the last decade [303, 304, 305, 306].
The canonical form of a discrete Hawkes process can be expressed as follows [277, 290, 291, 292, 294, 295, 302, 307]

$$\lambda(t) = \mu(t) + \sum_{t_i < t} \nu(t - t_i),$$  \hspace{1cm} (1.26)

where $\lambda(t) > 0$ specifies the rate of a terrorist attack at time $t$, $\mu(t) > 0$ is the background rate, and the response function $\nu(s) > 0$ defines the elevated rate in response to an event at a prior time $t_i$. The quantity $\lambda(t) - \mu(t)$ is the self-excitation contribution to the terrorist attack rate. Typically $\nu(s)$ is a decreasing function of $t - t_i$, so that the enhanced likelihood decays as $t_i$ moves into a more distant past. While other forms of decay functions have been adopted in the literature [308, 309, 310], a common choice is the exponential form [277, 278, 294, 295, 307, 310, 311]

$$\nu(s) = k_0 \omega \exp(-\omega s),$$  \hspace{1cm} (1.27)

where $\omega > 0$ specifies how fast the self-exciting effect dissipates over time, and $k_0 = \int_0^\infty \nu(s) \, ds$ represents the increased attack rate immediately after a prior attack. Eqs. 1.26–1.27 with a constant $\mu(t) \equiv \mu_0$ represent the most basic version of a self-exciting model for terrorist attacks, consisting of only three parameters: $\mu_0$, $k_0$, and $\omega$. For a given time series data $0 \leq t_i \leq T$ for $i = 1, 2, 3, \ldots, n$, parameters can be fitted through the methods of maximum likelihood estimation (MLE), which finds a set of parameter values that maximizes the following log-likelihood function $L$ [312, 313, 314]

$$\log L = \sum_{i=1}^n \log (\lambda(t_i)) - \int_0^T \lambda(t) \, dt.$$  \hspace{1cm} (1.28)

The first term on the right hand side of Eq. 1.28 corresponds to the likelihood of observing a data set given a probability distribution $\lambda(t)$, while the second term accounts for the constraint $\int_0^T \lambda(t) \, dt = 1$. White et al. [309] interpret the three parameters that appear in Eq. 1.26 respectively as the “intrinsic risk” ($\mu_0$), “volatility” ($k_0$), and “resilience” ($\omega^{-1}$) of terrorist activities within a country. The same authors apply the self-exciting model in Eqs. 1.26–1.27 to three southeast Asian countries – Indonesia, Philippines, and Thailand – finding significant variance among them, consistent with their respective sociopolitical contexts. This study demonstrates that the simple self-exciting model can be useful in quantitatively assessing terrorist threats in different geopolitical regions. Regional variations were also found in the underlying mechanisms of daily death tolls in four distinct Iraqi districts [277], and between Syria and England [315]. Najaf, the relatively more quiescent Iraqi region among the four districts examined in [277], displays the lowest intensity and shortest decay time of self-excitation. Syria exhibits highly clustered, self-excited civilian deaths, while England mainly displays only a background death rate [315]. Regional variations in the self-exciting parameters $\mu_0$, $k_0$ and $\omega$ due to local sociopolitics imply that these parameters may also change temporally. This is shown in [278] where the dynamics of improvised explosive device (IED) attacks in Northern Ireland was analyzed. The history of the Provisional Irish Republican Army (PIRA) was divided
Fig. 1.12 Panel (a) shows the histogram of number of fatalities per day in Fallujah, Iraq from March 20th 2003 to December 31st 2007. Note the emergence of three peaks. Panel (b) shows the total $\lambda(t)$, and time-dependent background $\mu(t)$, rates fitted from the data in panel (a). Their difference, $\lambda(t) - \mu(t)$, represents the self-excitation component of the terrorist attack rate. It is interesting to note that the first peak, around day 500, is mostly attributed to self-excitation, whereas the second and third peaks, around days 1250 and 1500 respectively, are driven primarily by rises to the background rate. Taken and modified from Ref. [277].

into five temporal phases, each defined by major events, and Eqs. 1.26–1.27 were applied to each. Significant changes in model parameters were found across the five time intervals: for example, when PIRA fractured into smaller cells in phase two, increased background rates, shortened time scales, and decreased volatility were observed.

While the sparsity of regional terrorism records often hinders the applicability of overly sophisticated models [308, 309], or yield only marginal gains, the basic Hawkes process still allows for the inclusion of time dependence in the three fundamental parameters, particularly in the background rate $\mu(t)$, which should naturally reflect local socio-political scenarios. Simple time-dependent functions for $\mu(t)$ include step functions, ramp functions, and heat kernels [277, 310, 316]. An example of the self-excitation model using a time-dependent $\mu(t)$ is displayed in Fig. 1.12, fitted for the daily death tolls in Fallujah, Iraq from March 20th 2003 to December 31st 2007, showing the respective contributions from the background rate $\mu(t)$, and the self-exciting process $\lambda(t) - \mu(t)$. Lewis et al. [277], Khraibani and Khraibani [310], and Johnson et al. [316] observe that given sufficient data, models with a constant $\mu(t) = \mu_0$ are either inadequate or consistently outperformed by models with time-dependent $\mu(t)$. For more flexible time dependence, a nonparametric method of Maximum Penalized Likelihood Estimation (MPLE) is proposed in [317]. Without assuming any specific functional form for $\mu(t)$ and $\nu(t)$, the method finds the best fitting functions via a log-likelihood function modified from Eq. 1.28

$$\log L = \sum_{i=1}^{n} \log (\lambda(t_i)) - \int_{0}^{T} \lambda(t)dt - \alpha_1 R(\mu) - \alpha_2 R(\nu),$$

(1.29)
where $R$ is a roughness penalty function (e.g., the $L^2$-norm of the first-order derivative of $\mu$ and $\nu$) which ensures certain regularities of the time-dependent functions $\mu$ and $\nu$, and $\alpha_{1,2}$ are the relative weights [318]. The basic self-exciting process in Eqs. 1.26–1.27 can also be extended to include mutual-excitation, whereby a distinct set of occurrences occurring at times $\{\tau_j\}$ further influence $\lambda(t)$. A possible representation is as follows [278]

$$\lambda(t) = \mu(t) + \sum_{i: t_i < t} \nu(t - t_i) + \sum_{j: \tau_j < t} \eta(t - \tau_j), \quad (1.30)$$

where the $\{\tau_j\}$ timings exert their influence on $\lambda(t)$ through the cross-response function $\eta(s)$. The original study on the distribution of IED attacks by PIRA [278] was revisited using Eq. 1.30 to examine whether attacks in different North Ireland counties and its capital Belfast could influence each other. From the analysis it was concluded that geographic mutual-excitation was far less influential than self-excitation. A similar conclusion was reached in independent studies of terrorist attacks in Iraq [311]. Here, spatial diffusion was included to the background rate $\mu(t)$, but the resulting analysis failed to capture the regional spread of terrorist activities. In addition to different geographic regions, the mutual-excitation model in Eq. 1.30 can also be used to examine interactions between different actors. For example, the IDE attacks by PIRA were further analyzed to verify whether the corresponding $\lambda(t)$ was affected by the actions of the British Security Force (BSF) [278]. The interesting finding was that PIRA's activities intensified much more as a reaction to Catholic civilian deaths caused by BSF, rather than to direct BSF killings of PIRA members. Such cross examinations may provide valuable insight for policy makers in devising guidelines for counter-terrorism operations. Independent work combined self-exciting processes with the hurdle model to analyze the daily number of terrorist attacks in Indonesia between 1994 and 2007 [308]. The hurdle model describes the frequency-position and the severity-height of events as two separate stochastic processes [319, 320, 321]. In the work of [308] the dates and number of terrorist attacks per day are chosen as the frequency and severity of the hurdles, respectively. Using a Riemann zeta distribution for the severity component, the authors found that incorporating self-exciting processes in the frequency component improves the fitting of the hurdle model to the data, and confirmed the hypothesis of a short-term increase of terrorism risk following an attack.

As for why terrorist activities exhibit self-exciting characteristics, some believe the reason is economic, as it may be more efficient in terms of costs and benefits to strike as many times as possible before any window of opportunity closes [322, 323]. Others point to copy-cat behaviors, as terrorists in the same organization or like-minded extremists tend to learn from each other. A positive-feedback may thus be generated through imitation, so that a single act can give rise to a cascade of events [321]. The relationship between the harboring of radical beliefs and the execution of terrorist attacks may be also addressed using hidden Markov models, where radical opinions can be regarded as a hidden layer of latent states, from which terrorist attacks manifest as observables [324, 325]. Regardless of the under-
lying reason, counter-terrorism intelligence may exploit the non-uniform likelihood of terrorist attacks to optimally allocate resources.

1.7 Summary and conclusions

In this brief review, we surveyed a few mathematical models that studied radicalization trajectories, how extremist groups organize, and patterns of terrorist attacks. These quantitative models are built from findings derived from the sociological literature, case studies, interviews and surveys, counter-terror reports, political science theories, data analysis and empirical observations, which are all limited by the covert nature of terrorist organizations and by the difficulty in conducting controlled tests. Much work is still necessary for a more complete understanding of the phenomena at hand, from all angles. For example, while data and interviews may offer an *a posteriori* understanding of why or how some individuals radicalize or become terrorists, they cannot explain why others exposed to the same environment do not. Most researchers show that early intervention is the optimal way to avoid the emergence of large scale radical factions and operative terrorists, but the legal implications of preemptive, invasive policy making must be factored in. Other advancements include a more rigorous comparative analysis to characterize terrorism in different historical and sociopolitical contexts, better distinction between online and offline radicalization, the emergence of lone wolves as opposed to group dynamics. Developing novel methods, including machine learning approaches, to dissect information from material that terrorists themselves divulge, in the form of videos and other internet content would offer better insight into their *modus pensandi*. Most work has focused on traditional attacks, but new forms of terrorism that would elicit different societal and governmental responses should also be investigated, such as the use of chemical, biological, radiological weapons. How to prevent the weaponization of trucks, airplanes, drinks or other materials is also of great concern. Exploring the role of the media, how to best use the internet as a counter-terrorism tool, how to create new channels of communications between counter-terror agencies across geographical boundaries, would help shape innovative intervention methods. Finally, finding potential pathways out of radicalization, and what to do with former foreign fighters returning to their home countries are also important issues.

Any attempt to answer these questions must involve experts and practitioners from many disciplines, from humanities and the social sciences to economics, computer and quantitative scientists. Mathematical modeling has great potential in this respect, as it allows to build predictive models where the consequences of different inputs, strategies, and actors may be examined, where costs and information constraints may be quantified. However, putative players, strategies and mechanisms must be cogently included. It is thus important that a common language between social and mathematical scientists be developed, that academic and practical opportunities for dialogue be created, and that each community actively engage with the
other. Great care and clarity must be exercised in translating any new finding into actual policy-making.

1.8 Acknowledgments

This work was supported by the ARO through grants W1911NF-14-1-0472 and W1911NF-16-1-0165; and by the NSF through grants DMS-1440415 and DMS-1814090.

References

58. Decker S.H., Pyrooz D.C.: I’m down for a Jihad: How 100 years of gang research can inform the study of terrorism, radicalization and extremism *Perspectives on terrorism* **9** 104–112 (2015)
71. LaFree G., Yang S.M., Crenshaw M.: Trajectories of terrorism: Attack patterns of foreign groups that have targeted the United States, 1970 to 2004 Criminology and Public Policy 8 445–473 (2009)


124. Likert R.: A technique for the measurement of attitudes *Archives of Psychology* 140 1–55 (1932)


141. Leslie P.H.: On the use of matrices in certain population mathematics *Biometrika* 33 183–212 (1945)
156. Fraser C., Gouve C., Billig M. Risky shifts, cautious shifts, and group polarization *Journal of Social Psychology* **10** 7–30 (1971)
236. Lapan H.E., Sandler T.: The political economy of terrorism: To bargain or not to bargain: That is the question AEA Papers and Proceedings 78 16–21 (1988)
258. Feng Q., Cai H., Chen Z., Zhao X., Chen Y.: Using game theory to optimize allocation of defensive resources to protect multiple chemical facilities in a city against terrorist attacks *Journal of Loss Prevention in the Process Industries* 43 614–628 (2016)
1 Mathematical Models of Radicalization and Terrorism


